(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 338 978 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
29.06.2011 Bulletin 2011/26

(21) Application number: **10184482.7**

(22) Date of filing: **30.10.2002**

(51) Int Cl.:
C12N 1/21 *(2006.01)* ***C12N 9/12*** *(2006.01)*
C12N 15/52 *(2006.01)* ***C12N 15/62*** *(2006.01)*

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

(30) Priority: **30.10.2001 US 335949 P**
11.04.2002 US 122706
23.05.2002 US 154515

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
02803298.5 / 1 451 293

(71) Applicant: **454 Life Sciences Corporation**
Branford CT 06405 (US)

(72) Inventors:
- **Srinivasan, Maithreyan**
Hamden, CT 06516 (US)
- **Reifler, Michael**
Hamden, CT 06514 (US)

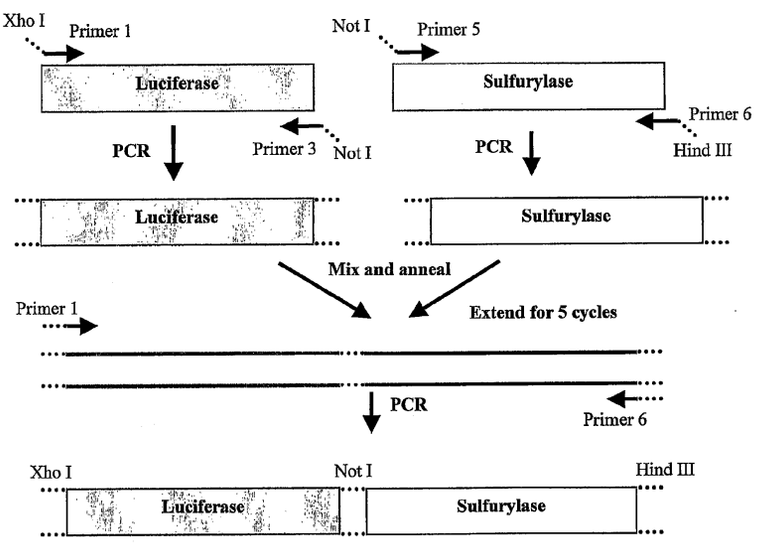
(74) Representative: **Finnie, Isobel Lara**
Mintz Levin Cohn Ferris Glovsky and Popeo Intellectual Property LLP
Alder Castle
10 Noble Street
London
EC2V 7JX (GB)

Remarks:
This application was filed on 30-09-2010 as a divisional application to the application mentioned under INID code 62.

(54) Novel sulfurylase-luciferase fusion proteins and thermostable sulfurylase

(57) The present invention relates to a thermostable ATP sulfurylase from B. Stearothermophilus. More specifically, this invention relates to fusion proteins comprising this ATP sulfurylase joined to a polypeptide that converts ATP into a detectable entity. Accordingly, this invention focuses on sulfurylase-luciferase fusion proteins.

FIGURE 1



EP 2 338 978 A1

Description

FIELD OF THE INVENTION

[0001] The invention relates generally to fusion proteins that are useful as reporter proteins, in particular to fusion proteins of ATP sulfurylase and luciferase which are utilized to achieve an efficient conversion of pyrophosphate (PPi) to light. This invention also relates to a novel thermostable sulfurylase which can be used in the detection of inorganic pyrophosphate, particularly in the sequencing of nucleic acid.

BACKGROUND OF THE INVENTION

[0002] ATP sulfurylase has been identified as being involved in sulfur metabolism. It catalyzes the initial reaction in the metabolism of inorganic sulfate (SO_4^{-2}); see e.g., Robbins and Lipmann, 1958. J. Biol. Chem. 233: 686-690; Hawes and Nicholas, 1973. Biochem. J. 133: 541-550). In this reaction SO_4^{-2} is activated to adenosine 5'-phosphosulfate (APS). ATP sulfurylase is also commonly used in pyrophosphate sequencing methods. In order to convert pyrophosphate (PPi) generated from the addition of dNMP to a growing DNA chain to light, PPi must first be converted to ATP by ATP sulfurylase.

[0003] ATP produced by an ATP sulfurylase can also be hydrolyzed using enzymatic reactions to generate light. Light-emitting chemical reactions (*ie.,* chemiluminescence) and biological reactions (*i.e.,* bioluminescence) are widely used in analytical biochemistry for sensitive measurements of various metabolites. In bioluminescent reactions, the chemical reaction that leads to the emission of light is enzyme-catalyzed. For example, the luciferin-luciferase system allows for specific assay of ATP. Thus, both ATP generating enzymes, such as ATP sulfurylase, and light emitting enzymes, such as luciferase, could be useful in a number of different assays for the detection and/or concentration of specific substances in fluids and gases. Since high physical and chemical stability is sometimes required for enzymes involved in sequencing reactions, a thermostable enzyme is desirable.

[0004] Because the product of the sulfurylase reaction is consumed by luciferase, proximity between these two enzymes by covalently linking the two enzymes in the form of a fusion protein would provide for a more efficient use of the substrate. Substrate channeling is a phenomenon in which substrates are efficiently delivered from enzyme to enzyme without equilibration with other pools of the same substrates. In effect, this creates local pools of metabolites at high concentrations relative to those found in other areas of the cell. Therefore, a fusion of an ATP generating polypeptide and an ATP converting peptide could benefit from the phenomenon of substrate channeling and would reduce production costs and increase the number of enzymatic reactions that occur during a given time period.

[0005] All patents and publications cited throughout the specification are hereby incorporated by reference into this specification in their entirety in order to more fully describe the state of the art to which this invention pertains.

SUMMARY OF THE INVENTION

[0006] The invention provides a fusion protein comprising an ATP generating polypeptide bound to a polypeptide which converts ATP into an entity which is detectable. In one aspect, the invention provides a fusion protein comprising a sulfurylase polypeptide bound to a luciferase polypeptide. This invention provides a nucleic acid that comprises an open reading frame that encodes a novel thermostable sulfurylase polypeptide. In a further aspect, the invention provides for a fusion protein comprising a thermostable sulfurylase joined to at least one affinity tag.

[0007] In another aspect, the invention provides a recombinant polynucleotide that comprises a coding sequence for a fusion protein having a sulfurylase poylpeptide sequence joined to a luciferase polypeptide sequence. In a further aspect, the invention provides an expression vector for expressing a fusion protein. The expression vector comprises a coding sequence for a fusion protein having: (i) a regulatory sequence, (ii) a first polypeptide sequence of an ATP generating polypeptide and (iii) a second polypeptide sequence that converts ATP to an entity which is detectable. In an additional embodiment, the fusion protein comprises a sulfurylase polypeptide and a luciferase polypeptide. In another aspect, the invention provides a transformed host cell which comprises the expression vector. In an additional aspect, the invention provides a fusion protein bound to a mobile support. The invention also includes a kit comprising a sulfurylase-luciferase fusion protein expression vector.

[0008] The invention also includes a method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) introducing the template nucleic acid polymer into a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added; (b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophos-

phate will be released; (c) separately recovering each of the feedstocks from the polymerization environment; and (d) measuring the amount of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein in each of the recovered feedstocks to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer. In one embodiment, the amount of inorganic pyrophosphate is measured by the steps of: (a) adding adenosine-5'-phosphosulfate to the feedstock; (b) combining the recovered feedstock containing adenosine-5'-phosphosulfate with an ATP generating polypeptide-ATP converting polypeptide fusion protein such that any inorganic pyrophosphate in the recovered feedstock and the adenosine-5'-phosphosulfate will react to the form ATP and sulfate; (c) combining the ATP and sulfate-containing feedstock with luciferin in the presence of oxygen such that the ATP is consumed to produced AMP, inorganic pyrophosphate, carbon dioxide and light; and (d) measuring the amount of light produced.

[0009] In another aspect, the invention includes a method wherein each feedstock comprises adenosine-5'-phosphosulfate and luciferin in addition to the selected nucleotide base, and the amount of inorganic pyrophosphate is determined by reacting the inorganic pyrophosphate feedstock with an ATP generating polypeptide-ATP converting polypeptide fusion protein thereby producing light in an amount proportional to the amount of inorganic pyrophosphate, and measuring the amount of light produced.

[0010] In another aspect, the invention provides a method for sequencing a nucleic acid, the method comprising: (a) providing one or more nucleic acid anchor primers; (b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm; (c) annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing an effective amount of a sequencing primer to one or more copies of said covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of a ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the sequence of the nucleic acid.

[0011] In one aspect, the invention provides a method for sequencing a nucleic acid, the method comprising: (a) providing at least one nucleic acid anchor primer; (b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites; (c) annealing a first amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing a second amount of a sequencing primer to one or more copies of the covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, when the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, to yield a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of a ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the sequence of the nucleic acid at each reaction site that contains a nucleic acid template.

[0012] In another aspect, the invention includes a method of determining the base sequence of a plurality of nucleotides on an array, the method comprising the steps of: (a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, (b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (c) determining whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands through detection of a sequencing byproduct with a ATP generating polypeptide-ATP converting polypeptide fusion protein, thus indicating that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and (d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and

(e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

[0013] In one aspect, the invention includes a method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) introducing a plurality of template nucleic acid polymers into a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, each reaction chamber having a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added; (b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released; (c) detecting the formation of inorganic pyrophosphate with an ATP generating polypeptide-ATP converting polypeptide fusion protein to determine the identify of each nucleotide in the complementary polymer and thus the sequence of the template polymer.

[0014] In one aspect, the invention provides a method of identifying the base in a target position in a DNA sequence of sample DNA including the steps comprising: (a) disposing sample DNA within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, said DNA being rendered single stranded either before or after being disposed in the reaction chambers, (b) providing an extension primer which hybridizes to said immobilized single-stranded DNA at a position immediately adjacent to said target position; (c) subjecting said immobilized single-stranded DNA to a polymerase reaction in the presence of a predetermined nucleotide triphosphate, wherein if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer then a sequencing reaction byproduct is formed; and (d) identifying the sequencing reaction byproduct with a ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the nucleotide complementary to the base at said target position.

[0015] The invention also includes a method of identifying a base at a target position in a sample DNA sequence comprising: (a) providing sample DNA disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, said DNA being rendered single stranded either before or after being disposed in the reaction chambers; (b) providing an extension primer which hybridizes to the sample DNA immediately adjacent to the target position; (c) subjecting the sample DNA sequence and the extension primer to a polymerase reaction in the presence of a nucleotide triphosphate whereby the nucleotide triphosphate will only become incorporated and release pyrophosphate (PPi) if it is complementary to the base in the target position, said nucleotide triphosphate being added either to separate aliquots of sample-primer mixture or successively to the same sample-primer mixture; (d) detecting the release of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein to indicate which nucleotide is incorporated.

[0016] In one aspect, the invention provides a method of identifying a base at a target position in a single-stranded sample DNA sequence, the method comprising: (a) providing an extension primer which hybridizes to sample DNA immediately adjacent to the target position, said sample DNA disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 um, said DNA being rendered single stranded either before or after being disposed in the reaction chambers; (b) subjecting the sample DNA and extension primer to a polymerase reaction in the presence of a predetermined deoxynucleotide or dideoxynucleotide whereby the deoxynucleotide or dideoxynucleotide will only become incorporated and release pyrophosphate (PPi) if it is complementary to the base in the target position, said predetermined deoxynucleotides or dideoxynucleotides being added either to separate aliquots of sample-primer mixture or successively to the same sample-primer mixture, (c) detecting any release of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein to indicate which deoxynucleotide or dideoxynucleotide is incorporated;characterized in that, the PPi-detection enzyme(s) are included in the polymerase reaction step and in that in place of deoxy- or dideoxy adenosine triphosphate (ATP) a dATP or ddATP analogue is used which is capable of acting as a substrate for a polymerase but incapable of acting as a substrate for a said PPi—detection enzyme.

[0017] In another aspect, the invention includes a method of determining the base sequence of a plurality of nucleotides on an array, the method comprising: (a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, (b) converting PPi into light with an ATP generating polypeptide-ATP converting polypeptide fusion protein; (c) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device; (d) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions; (e) determining a light intensity for each of said discrete regions from the corresponding electrical signal; (f) recording the variations of said electrical signals with time.

[0018] In one aspect, the invention provides a method for sequencing a nucleic acid, the method comprising:(a) providing one or more nucleic acid anchor primers; (b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber,

wherein the reaction chambers have a center to center spacing of between 5 to 200 μm;(c) converting PPi into a detectable entity with the use of an ATP generating polypeptide-ATP converting polypeptide fusion protein; (d) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device; (e) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions; (f) determining a light intensity for each of said discrete regions from the corresponding electrical signal; (g) recording the variations of said electrical signals with time.

[0019] In another aspect, the invention includes a method for sequencing a nucleic acid, the method comprising: (a) providing at least one nucleic acid anchor primer; (b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites; (c) converting PPi into a detectable entity with an ATP generating polypeptide-ATP converting polypeptide fusion protein; (d) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device; (e) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions; (f) determining a light intensity for each of said discrete regions from the corresponding electrical signal; (g) recording the variations of said electrical signals with time.

[0020] In another aspect, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of an amino acid sequence of SEQ ID NO: 2; (b) a variant of a mature form of an amino acid sequence of SEQ ID NO: 2; an amino acid sequence of SEQ ID NO: 2; (c) a variant of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 5% of amino acid residues from said amino acid sequence; (d) and at least one conservative amino acid substitution to the amino acid sequences in (a), (b), (c) or (d). The invention also includes an antibody that binds immunospecifically to the polypeptide of (a), (b), (c) or (d).

[0021] In another aspect, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of an amino acid sequence of SEQ ID NO: 2; (b) a variant of a mature form of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 5% of the amino acid residues from the amino acid sequence of said mature form; (c) an amino acid sequence of SEQ ID NO: 2; (d) a variant of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence; a nucleic acid fragment encoding at least a portion of a polypeptide comprising an amino acid sequence of SEQ ID NO: 2, or a variant of said polypeptide, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 5% of amino acid residues from said amino acid sequence; (e) and a nucleic acid molecule comprising the complement of (a), (b), (c), (d) or (e). In a further aspect, the invention provides a nucleic acid molecule wherein the nucleic acid molecule comprises nucleotide sequence selected from the group consisting of: (a) a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 20% of the nucleotides in the coding sequence in said first nucleotide sequence differ from said coding sequence; an isolated second polynucleotide that is a complement of the first polynucleotide; (b) and a nucleic acid fragment of (a) or (b). The invention also includes a vector comprising the nucleic acid molecule of (a) or (b). In another aspect, the invention includes a cell comprising the vector.

[0022] In a further aspect, the invention includes a method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) introducing the template nucleic acid polymer into a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added; (b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released; (c) separately recovering each of the feedstocks from the polymerization environment; and (d) measuring the amount of PPi with an ATP sulfurylase and a luciferase in each of the recovered feedstocks to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer.

[0023] In another aspect, the invention provides a method for sequencing a nucleic acid, the method comprising: (a) providing one or more nucleic acid anchor primers; (b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities in an array on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm and at least 400,000 discrete sites; (c) annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing an effective amount of a sequencing

primer to one or more copies of said covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of an ATP sulfurylase and a luciferase, thereby determining the sequence of the nucleic acid.

[0024] In another aspect, the invention provides a method for sequencing a nucleic acid, the method comprising: (a) providing at least one nucleic acid anchor primer; (b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites; (c) annealing a first amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing a second amount of a sequencing primer to one or more copies of the covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, when the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, to yield a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of a thermostable sulfurylase and a luciferase, thereby determining the sequence of the nucleic acid at each reaction site that contains a nucleic acid template.

[0025] In a further aspect, the invention includes a method of determining the base sequence of a plurality of nucleotides on an array, the method comprising: (a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, (b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (c) detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands through detection of a sequencing byproduct with a thermostable sulfurylase and luciferase, thus indicating that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and (d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and (e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

BRIEF DESCRIPTION OF THE DRAWINGS

[0026]

FIG. 1 is one embodiment for a cloning strategy for obtaining the luciferase-sulfurylase sequence.

FIG. 2A and 2B show the preparative agarose gel of luciferase and sulfurylase as well as sulfurylase-luciferase fusion genes.

FIG. 3 shows the results of experiments to determine the activity of the luciferase-sulfurylase fusion protein on NTA-agarose and MPG-SA solid supports.

DETAILED DESCRIPTION OF THE INVENTION

[0027] This invention provides a fusion protein containing an ATP generating polypeptide bound to a polypeptide which converts ATP into an entity which is detectable. As used herein, the term "fusion protein" refers to a chimeric protein containing an exogenous protein fragment joined to another exogenous protein fragment. The fusion protein could include an affinity tag to allow attachment of the protein to a solid support or to allow for purification of the recombinant fusion protein from the host cell or culture supernatant, or both.

[0028] In a preferred embodiment, the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote. The eukaryote could be an animal, plant, fungus or yeast. In some embodiments, the animal is a mammal, rodent, insect, worm, mollusk, reptile, bird and amphibian. Plant sources of the polypeptides include but are not limited to *Arabidopsis thaliana, Brassica napus, Allium sativum, Amaranthus caudatus, Hevea brasiliensis, Hordeum vulgare, Lycopersicon esculentum, Nicotiana tabacum, Oryza sativum, Pisum sativum, Populus trichocarpa, Solanum tuberosum, Secale cereale, Sambucus nigra, Ulmus americana* or *Triticum aestivum.* Examples of fungi

include but are not limited to *Penicillum chrysogenum, Stachybotrys chartarum, Aspergillus fumigatus, Podospora anserina* and *Trichoderma reesei.* Examples of sources of yeast include but are not limited to *Saccharomyces cerevisiae, Candida tropicalis, Candida lypolitica, Candida utilis, Kluyveromyces lactis, Schizosaccharomyces pombe, Yarrowia lipolytica, Candida spp., Pichia spp.* and *Hansenula spp.*.

[0029] The prokaryote source could be bacteria or archaea. In some embodiments, the bacteria is E. *coli, B. subtilis, Streptococcus gordonii,* flavobacteria or green sulfur bacteria. In other embodiments, the archaea is *Sulfolobus, Thermococcus, Methanobacterium, Halococcus, Halobacterium* or *Methanococcus jannaschii.*

[0030] The ATP generating polypeptide can be a ATP sulfurylase, hydrolase or an ATP synthase. In a preferred embodiment, the ATP generating polypeptide is ATP sulfurylase. In one embodiment, the ATP sulfurylase is a thermostable sulfurylase cloned from *Bacillus stearothermophilus* (Bst) and comprising the nucleotide sequence of SEQ ID NO:1. This putative gene was cloned using genomic DNA acquired from ATCC (Cat. No. 12980D). The gene is shown to code for a functional ATP sulfurylase that can be expressed as a fusion protein with an affinity tag. The disclosed Bst sulfurylase nucleic acid (SEQ ID NO:1) includes the 1247 nucleotide sequence. An open reading frame (ORF) for the mature protein was identified beginning with an ATG codon at nucleotides 1-3 and ending with a TAA codon at nucleotides 1159-1161. The start and stop codons of the open reading frame are highlighted in bold type. The putative untranslated regions are underlined and found upstream of the initiation codon and downstream from the termination codon.

Bst Thermostable Sulfurylase Nucleotide Sequence (SEQ ID NO:1)

GTTATGAACATGAGTTTGAGCATTCCGCATGGCGGCACATTGATCAACCGTTGGAATCCG 60
GATTACCCAATCGATGAAGCAACGAAAACGATCGAGCTGTCCAAAGCCGAACTAAGCGAC 120
CTTGAGCTGATCGGCACAGGCGCCTACAGCCCGCTCACCGGGTTTTTAACGAAAGCCGAT 180
TACGATGCGGTCGTAGAAACGATGCGCCTCGCTGATGGCACTGTCTGGAGCATTCCGATC 240
ACGCTGGCGGTGACGGAAGAAAAAGCGAGTGAACTCACTGTCGGCGACAAAGCGAAACTC 300
GTTTATGGCGGCGACGTCTACGGCGTCATTGAAATCGCCGATATTTACCGCCCGGATAAA 360
ACGAAAGAAGCCAAGCTCGTCTATAAAACCGATGAACTCGCTCACCCGGGCGTGCGCAAG 420
CTGTTTGAAAAACCAGATGTGTACGTCGGCGGAGCGGTTACGCTCGTCAAACGGACCGAC 480
AAAGGCCAGTTTGCTCCGTTTTATTTCGATCCGGCCGAAACGCGGAAACGATTTGCCGAA 540
CTCGGCTGGAATACCGTCGTCGGCTTCCAAACACGCAACCCGGTTCACCGCGCCCATGAA 600
TACATTCAAAAATGCGCGCTTGAAATCGTGGACGGCTTGTTTTTAAACCCGCTCGTCGGC 660
GAAACGAAAGCGGACGATATTCCGGCCGACATCCGGATGGAAAGCTATCAAGTGCTGCTG 720
GAAAACTATTATCCGAAAGACCGCGTTTTCTTGGGCGTCTTCCAAGCTGCGATGCGCTAT 780
GCCGGTCCGCGCGAAGCGATTTTCCATGCCATGGTGCGGAAAAACTTCGGCTGCACGCAC 840
TTCATCGTCGGCCGCGACCATGCGGGCGTCGGCAACTATTACGGCACGTATGATGCGCAA 900
AAAATCTTCTCGAACTTTACAGCCGAAGAGCTTGGCATTACACCGCTCTTTTTCGAACAC 960
AGCTTTTATTGCACGAAATGCGAAGGCATGGCATCGACGAAAACATGCCCGCACGACGCA 1020
CAATATCACGTTGTCCTTTCTGGCACGAAAGTCCGTGAAATGTTGCGTAACGGCCAAGTG 1080
CCGCCGAGCACATTCAGCCGTCCGGAAGTGGCCGCCGTTTTGATCAAAGGGCTGCAAGAA 1140
CGCGAAACGGTCACCCCGTCGACACGCTAAAGGAGGAGCGAGATGAGCACGAATATCGTT 1200
TGGCATCATACATCGGTGACAAAAGAAGATCGCCGCCAACGCAACGG 1247

[0031] The Bst sulfurylase polypeptide (SEQ ID NO:2) is 386 amino acid residues in length and is presented using the three letter amino acid code.

Bst Sulfurylase Amino Acid Sequence (SEQ ID NO:2)

Met Ser Leu Ser Ile Pro His Gly Gly Thr Leu Ile Asn Arg Trp Asn
1 5 10 15

Pro Asp Tyr Pro Ile Asp Glu Ala Thr Lys Thr Ile Glu Leu Ser Lys
20 25 30

Ala Glu Leu Ser Asp Leu Glu Leu Ile Gly Thr Gly Ala Tyr Ser Pro
35 40 45

Leu Thr Gly Phe Leu Thr Lys Ala Asp Tyr Asp Ala Val Val Glu Thr

50 55 60

Met Arg Leu Ala Asp Gly Thr Val Trp Ser Ile Pro Ile Thr Leu Ala
65 70 75

Val Thr Glu Glu Lys Ala Ser Glu Leu Thr Val Gly Asp Lys Ala Lys
80 85 90 95

Leu Val Tyr Gly Gly Asp Val Tyr Gly Val Ile Glu Ile Ala Asp Ile
100 105 110

Tyr Arg Pro Asp Lys Thr Lys Glu Ala Lys Leu Val Tyr Lys Thr Asp
115 120 125

Glu Leu Ala His Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val
130 135 140

Tyr Val Gly Gly Ala Val Thr Leu Val Lys Arg Thr Asp Lys Gly Gln
145 150 155
Phe Ala Pro Phe Tyr Phe Asp Pro Ala Glu Thr Arg Lys Arg Phe Ala
160 165 170 175

Glu Leu Gly Trp Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val
180 185 190

His Arg Ala His Glu Tyr Ile Gln Lys Cys Ala Leu Glu Ile Val Asp
195 200 205

Gly Leu Phe Leu Asn Pro Leu Val Gly Glu Thr Lys Ala Asp Asp Ile
210 215 220

Pro Ala Asp Ile Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr
225 230 235

Tyr Pro Lys Asp Arg Val Phe Leu Gly Val Phe Gln Ala Ala Met Arg
240 245 250 255

Tyr Ala Gly Pro Arg Glu Ala Ile Phe His Ala Met Val Arg Lys Asn
260 265 270

Phe Gly Cys Thr His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly
275 280 285

Asn Tyr Tyr Gly Thr Tyr Asp Ala Gln Lys Ile Phe Ser Asn Phe Thr
290 295 300

Ala Glu Glu Leu Gly Ile Thr Pro Leu Phe Phe Glu His Ser Phe Tyr
305 310 315

Cys Thr Lys Cys Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp
320 325 330 335

Ala Gln Tyr His Val Val Leu Ser Gly Thr Lys Val Arg Glu Met Leu
340 345 350

Arg Asn Gly Gln Val Pro Pro Ser Thr Phe Ser Arg Pro Glu Val Ala
355 360 365

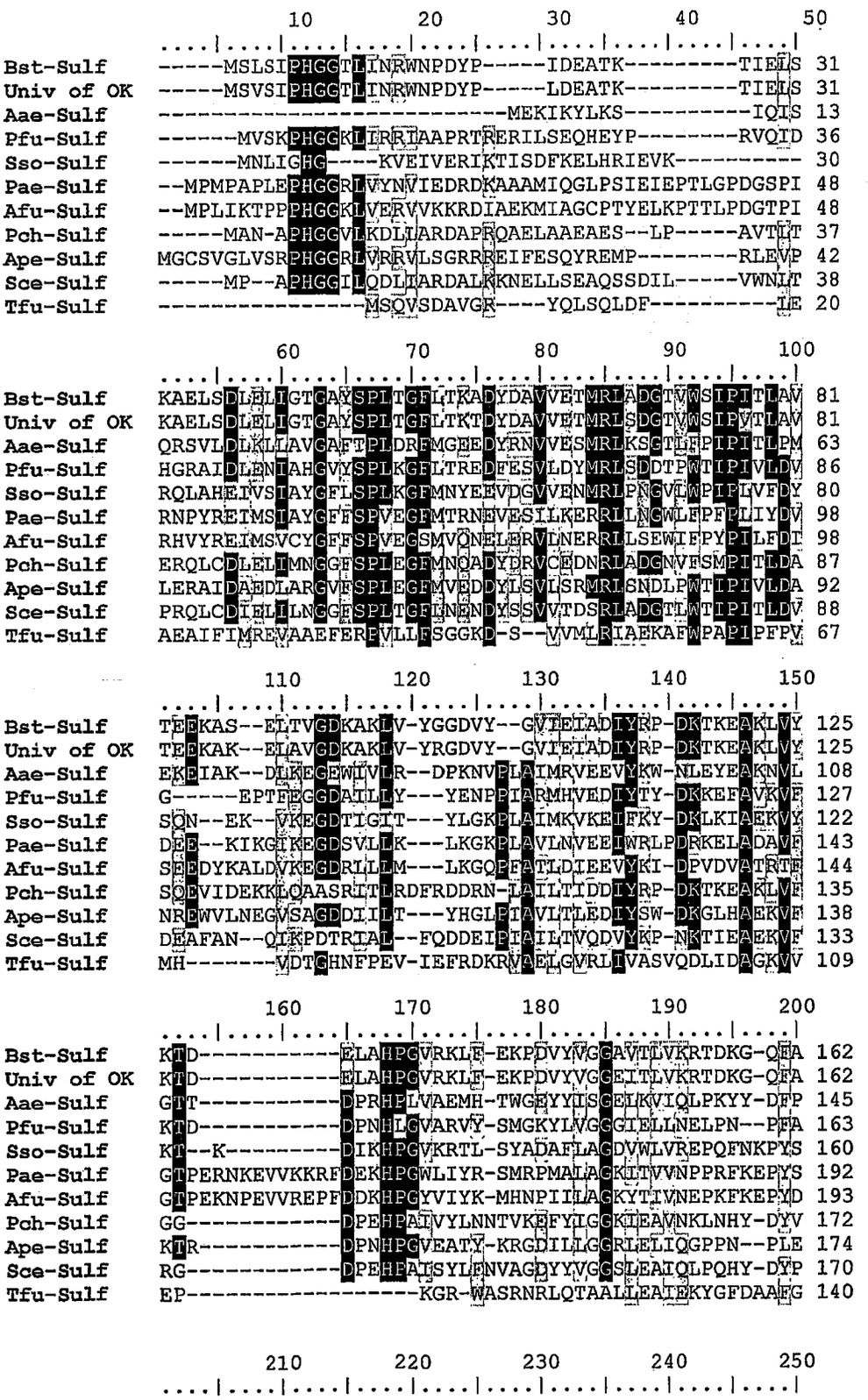
Ala Val Leu Ile Lys Gly Leu Gln Glu Arg Glu Thr Val Thr Pro Ser
370 375 380

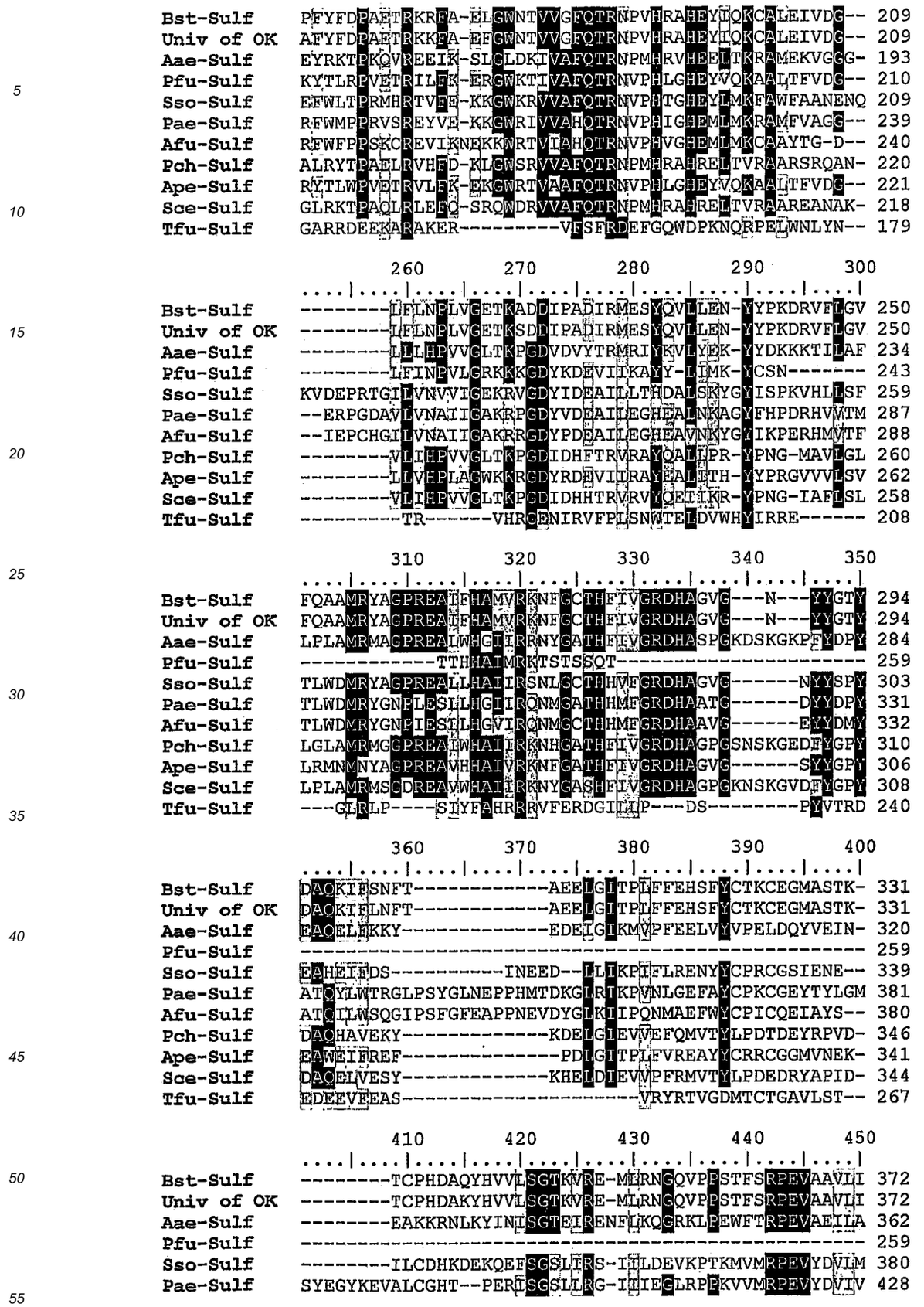
Thr Arg
385

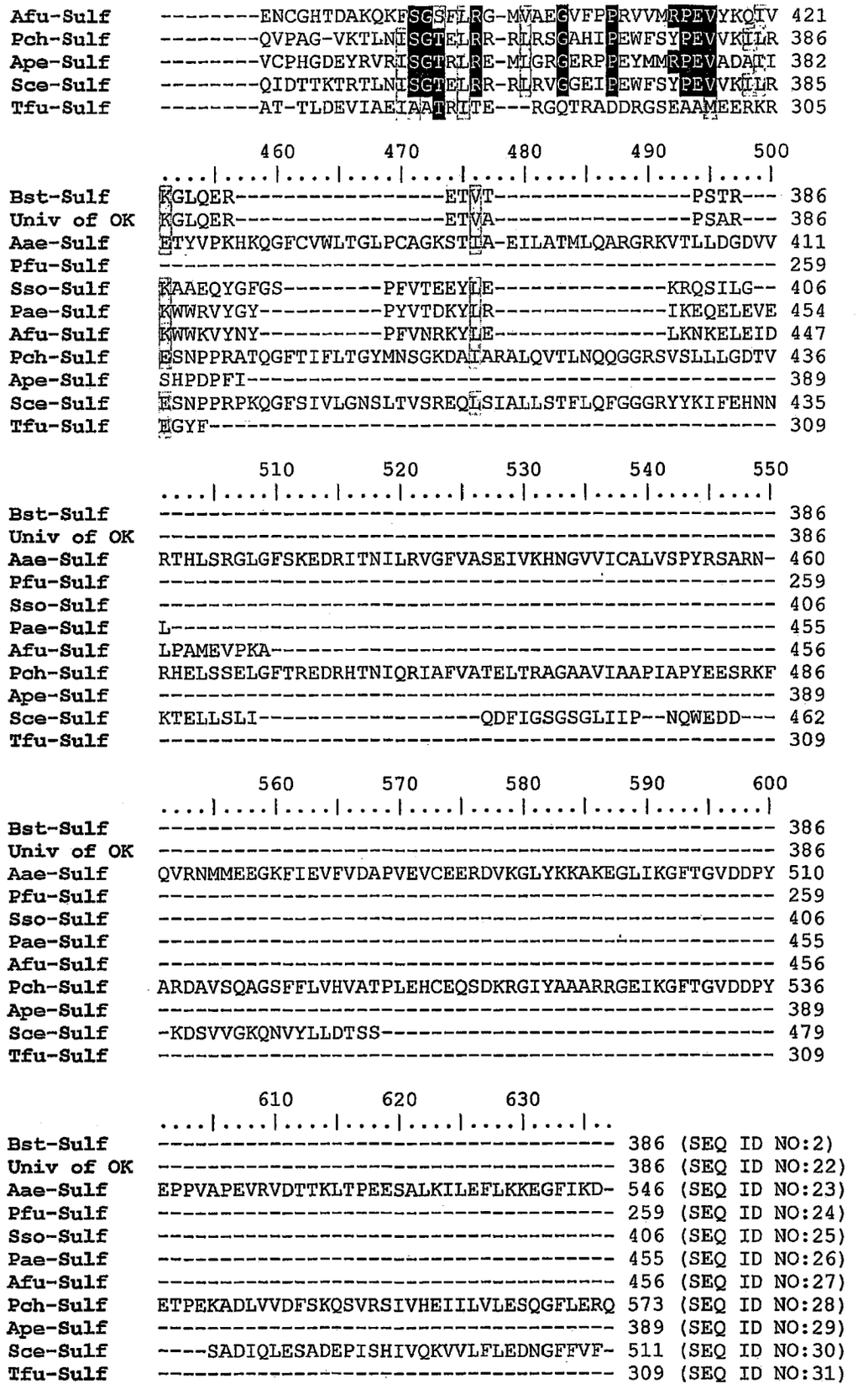
[0032] In one embodiment, the thermostable sulfurylase is active at temperatures above ambient to at least 50˚C. This property is beneficial so that the sulfurylase will not be denatured at higher temperatures commonly utilized in polymerase chain reaction (PCR) reactions or sequencing reactions. In one embodiment, the ATP sulfurylase is from a thermophile. The thermostable sulfurylase can come from thermophilic bacteria, including but not limited to, *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

[0033] The homology of twelve ATP sulfurylases can be shown graphically in the ClustalW analysis in Table1. The alignment is of ATP sulfurylases from the following species: *Bacillus stearothermophilus* (Bst), University of Oklahoma —Strain 10 (Univ of OK), *Aquifex aeolicus* (Aae), *Pyrococcus furiosus* (Pfu), *Sulfolobus solfataricus* (Sso), *Pyrobaculum aerophilum* (Pae), *Archaeoglobus fulgidus* (Afu), *Penicillium chrysogenum (Pch)*,*Aeropyrum pernix* (Ape), *Saccharomyces cerevisiae* (Sce), and *Thermomonospora fusca* (Tfu). **Table 1: ClustalW Analysis of ATP Sulfurylase Amino Acid Sequence**

Table 1: ClustalW Analysis of ATP Sulfurylase Amino Acid Sequence




Bst-Sulf PFYFDPAETRKRFA-ELGWNTVVGFQTRNPVHRAHEYIQKCALEIVDG-- 209
Univ of OK AFYFDPAETRKKFA-EFGWNTVVGFQTRNPVHRAHEYIQKCALEIVDG-- 209
Aae-Sulf EYRKTPKQVREEIK-SLGLDKIVAFQTRNPMHRVHEELTKRAMEKVGGG- 193
Pfu-Sulf KYTLRPVETRILEK-ERGWKTIVAFQTRNVPHLGHEYVQKAALTFVDG-- 210
Sso-Sulf EFWLTPRMHRTVEE-KKGWKRVVAFQTRNVPHTGHEYLMKFAWFAANENQ 209
Pae-Sulf RFWMPPRVSREYVE-KKGWRIVVAHQTRNVPHIGHEMLMKRAMFVAGG-- 239
Afu-Sulf RFWFPPSKCREVIKNEKKWRTVIAHQTRNVPHVGHEMLMKCAAYTG-D-- 240
Pch-Sulf ALRYTPAELRVHFD-KLGWSRVVAFQTRNPMHRAHRELTVRAARSRQAN- 220
Ape-Sulf RYTLWPVETRVLEK-EKGWRTVAAFQTRNVPHLGHEYVQKAALTFVDG-- 221
Sce-Sulf GLRKTPAQLRLEEQ-SRQWDRVVAFQTRNPMHRAHRELTVRAAREANAK- 218
Tfu-Sulf GARRDEEKARAKER---------VFSFRDEFGQWDPKNQRPELWNLYN-- 179
260 270 280 290 300
Bst-Sulf --------LFLNPLVGETKADDIPADIRMESYQVLLEN-YYPKDRVFLGV 250
Univ of OK --------LFLNPLVGETKSDDIPADIRMESYQVLLEN-YYPKDRVFLGV 250
Aae-Sulf --------LLLHPVVGLTKPGDVDVYTRMRIYKVLYEK-YYDKKKTILAF 234
Pfu-Sulf --------LFINPVLGRKKKGDYKDEVIIKAYY-LIMK-YCSN-------- 243
Sso-Sulf KVDEPRTGILVNVVIGEKRVGDYIDEAILLTHDALSKYGYISPKVHLLSF 259
Pae-Sulf --ERPGDAVLVNAIIGAKRPGDYVDEAILEGHEALNKAGYFHPDRHVVTM 287
Afu-Sulf --IEPCHGILVNAIIGAKRRGDYPDEAILEGHEAVNKYGYIKPERHMVTF 288
Pch-Sulf --------VLIHPVVGLTKPGDIDHFTRVRAYQALIPR-YPNG-MAVLGL 260
Ape-Sulf --------LLVHPLAGWKKRGDYRDEVIIRAYEALITH-YYPRGVVVLSV 262
Sce-Sulf --------VLIHPVVGLTKPGDIDHHTRVRVYQETIKR-YPNG-IAFLSL 258
Tfu-Sulf ----------TR------VHRGENIRVFPLSNWTELDVWHYIRRE------ 208
310 320 330 340 350
Bst-Sulf FQAAMRYAGPREATFHAMVRKNFGCTHFIVGRDHAGVG---N---YYGTY 294
Univ of OK FQAAMRYAGPREATFHAMVRKNFGCTHFIVGRDHAGVG---N---YYGTY 294
Aae-Sulf LPLAMRMAGPREALWHGIIRRNYGATHFIVGRDHASPGKDSKGKPFYDPY 284
Pfu-Sulf ------------TTHHAIMRKTSTSSQT---------- 259
Sso-Sulf TLWDMRYAGPREALLHAIIRSNLGCTHHVFGRDHAGVG------NYYSPY 303
Pae-Sulf TLWDMRYGNFLESILHGIIRQNMGATHHMFGRDHAATG------DYYDPY 331
Afu-Sulf TLWDMRYGNFIESILHGVIRQNMGCTHHMFGRDHAAVG------EYYDMY 332
Pch-Sulf LGLAMRMGGPREAIWHAIIRKNHGATHFIVGRDHAGPGSNSKGEDFYGPY 310
Ape-Sulf LRMNMNYAGPREAVHHAIVRKNFGATHFIVGRDHAGVG------SYYGPT 306
Sce-Sulf LPLAMRMSGDREAVWHAIIRKNYGASHFIVGRDHAGPGKNSKGVDFYGPY 308
Tfu-Sulf ---GLRLP----SIYFAHRRRVFERDGILIP---DS--------PYVTRD 240
360 370 380 390 400
Bst-Sulf DAQKIFSNFT----------AEELGITPLFFEHSFYCTKCEGMASTK- 331
Univ of OK DAQKIFLNFT----------AEELGITPLFFEHSFYCTKCEGMASTK- 331
Aae-Sulf EAQELFKKY----------EDEIGIKMVPFEELVYVPELDQYVEIN- 320
Pfu-Sulf ---------- 259
Sso-Sulf EAHEIFDS----------INEED--LLIKPIFLRENYYCPRCGSIENE-- 339
Pae-Sulf ATQYLWTRGLPSYGLNEPPHMTDKGLRTKPVNLGEFAYCPKCGEYTYLGM 381
Afu-Sulf ATQILWSQGIPSFGFEAPPNEVDYGLKIIPQNMAEFWYCPICQEIAYS-- 380
Pch-Sulf DAQHAVEKY----------KDELGIEVVEFQMVTYLPDTDEYRPVD- 346
Ape-Sulf EAWEIFREF----------PDIGITPLFVREAYYCRRCGGMVNEK- 341
Sce-Sulf DAQELVESY----------KHELDIEVVPFRMVTYLPDEDRYAPID- 344
Tfu-Sulf EDEEVEEAS----------VRYRTVGDMTCTGAVLST-- 267
410 420 430 440 450
Bst-Sulf --------TCPHDAQYHVVLSGTKVRE-MIRNGQVPPSTFSRPEVAAVLI 372
Univ of OK --------TCPHDAKYHVVLSGTKVRE-MLRNGQVPPSTFSRPEVAAVLI 372
Aae-Sulf --------EAKKRNLKYINISGTEIRENFLKQGRKLPEWFTRPEVAEILA 362
Pfu-Sulf ---------- 259
Sso-Sulf --------ILCDHKDEKQEFSGSLIRS-ILLDEVKPTKMVMRPEVYDVLM 380
Pae-Sulf SYEGYKEVALCGHT--PERISGSLLRG-IIIEGLRPEKVVMRPEVYDVIV 428



[0034] A thermostable sulfurylase polypeptide is encoded by the open reading frame ("ORF") of a thermostable sulfurylase nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a

polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, e.g., a stretch of DNA that would encode a protein of 50 amino acids or more.

[0035] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO:1 due to degeneracy of the genetic code and thus encode the same thermostable sulfurylase proteins as that encoded by the nucleotide sequences shown in SEQ ID NO: 1. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2. In addition to the thermostable sulfurylase nucleotide sequence shown in SEQ ID NO: 1 it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the thermostable sulfurylase polypeptides may exist within a population (e.g., the bacterial population). Such genetic polymorphism in the thermostable sulfurylase genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a thermostable sulfurylase protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the thermostable sulfurylase genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the thermostable sulfurylase polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the thermostable sulfurylase polypeptides, are intended to be within the scope of the invention.

[0036] Moreover, nucleic acid molecules encoding thermostable sulfurylase proteins from other species, and thus that have a nucleotide sequence that differs from the sequence SEQ ID NO:1 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the thermostable sulfurylase cDNAs of the invention can be isolated based on their homology to the thermostable sulfurylase nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. The invention further includes the nucleic acid sequence of SEQ ID NO: 1 and mature and variant forms thereof, wherein a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 11 % of the nucleotides in the coding sequence differ from the coding sequence.

[0037] Another aspect of the invention pertains to nucleic acid molecules encoding a thermostable sulfurylase protein that contains changes in amino acid residues that are not essential for activity. Such thermostable sulfurylase proteins differ in amino acid sequence from SEQ ID NO:2 yet retain biological activity. In separate embodiments, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 96%, 97%, 98% or 99% homologous to the amino acid sequence of SEQ ID NO:2. An isolated nucleic acid molecule encoding a thermostable sulfurylase protein homologous to the protein of SEQ ID NO: 2 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

[0038] Mutations can be introduced into SEQ ID NO:2 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains *(e.g.*, lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains *(e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains *(e.g.,* threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the thermostable sulfurylase protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a thermostable sulfurylase coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for thermostable sulfurylase biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

[0039] The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single

letter amino acid code.

[0040] The thermostable sulfurylase nucleic acid of the invention includes the nucleic acid whose sequence is provided herein, or fragments thereof. The invention also includes mutant or variant nucleic acids any of whose bases may be changed from the corresponding base shown herein while still encoding a protein that maintains its sulfurylase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

[0041] A thermostable sulfurylase nucleic acid can encode a mature thermostable sulfurylase polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

[0042] The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid *(i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated thermostable sulfurylase nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (e.g., brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

[0043] A nucleic acid molecule of the invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:1 as a hybridization probe, thermostable sulfurylase molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, et al., (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, et al., (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY,1993.)

[0044] A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to thermostable sulfurylase nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

[0045] As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

[0046] Fragments provided herein are defmed as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific

recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

[0047] Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 89% identity over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See e.g. Ausubel, et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY,1993, and below.

[0048] A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of thermostable sulfurylase polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a thermostable sulfurylase polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, e.g., frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions in SEQ ID NO: 1, as well as a polypeptide possessing thermostable sulfurylase biological activity. Various biological activities of the thermostable sulfurylase proteins are described below.

[0049] The thermostable sulfurylase proteins of the invention include the sulfurylase protein whose sequence is provided herein. The invention also includes mutant or variant proteins any of whose residues may be changed from the corresponding residue shown herein while still encoding a protein that maintains its sulfurylase-like activities and physiological functions, or a functional fragment thereof. The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention. This invention also includes a variant or a mature form of the amino acid sequence of SEQ ID NO:2, wherein one or more amino acid residues in the variant differs in no more than 4% of the amino acic residues from the amino acid sequence of the mature form.

[0050] Several assays have been developed for detection of the forward ATP sulfurylase reaction. The colorimetric molybdolysis assay is based on phosphate detection (see e.g., Wilson and Bandurski, 1958. J. Biol. Chem. 233: 975-981), whereas the continuous spectrophotometric molybdolysis assay is based upon the detection of NADH oxidation (see *e.g.*, Seubert, et al., 1983. Arch. Biochem. Biophys. 225: 679-691; Seubert, et al., 1985. Arch. Biochem. Biophys. 240: 509-523). The later assay requires the presence of several detection enzymes.

[0051] Suitable enzymes for converting ATP into light include luciferases, *e.g.*, insect luciferases. Luciferases produce light as an end-product of catalysis. The best known light-emitting enzyme is that of the firefly, *Photinus pyralis (Coleoptera).* The corresponding gene has been cloned and expressed in bacteria (see *e.g.,* de Wet, et al., 1985. Proc. Natl. Acad. Sci. USA 80: 7870-7873) and plants (see *e.g.,* Ow, et al., 1986. Science 234: 856-859), as well as in insect (see *e.g.,* Jha, et al., 1990. FEBS Lett. 274: 24-26) and mammalian cells (see *e.g.,* de Wet, et al., 1987. Mol. Cell. Biol. 7: 725-7373; Keller, et al., 1987. Proc. Natl. Acad. Sci. USA 82: 3264-3268). In addition, a number of luciferase genes from the Jamaican click beetle, *Pyroplorus plagiophihalamus (Coleoptera),* have recently been cloned and partially characterized (see *e.g.*, Wood, et al., 1989.J. Biolumin. Chemilumin. 4: 289-301; Wood, et al., 1989. Science 244: 700-702). Distinct luciferases can sometimes produce light of different wavelengths, which may enable simultaneous monitoring of light emissions at different wavelengths. Accordingly, these aforementioned characteristics are unique, and add new dimensions with respect to the utilization of current reporter systems.

[0052] Firefly luciferase catalyzes bioluminescence in the presence of luciferin, adenosine 5'-triphosphate (ATP), magnesium ions, and oxygen, resulting in a quantum yield of 0.88 (see *e.g.*, McElroy and Selinger, 1960. Arch. Biochem. Biophys. 88: 136-145). The firefly luciferase bioluminescent reaction can be utilized as an assay for the detection of ATP with a detection limit of approximately $1x10^{-13}$ M (see e.g., Leach, 1981.J. Appl. Biochem. 3: 473-517). In addition, the overall degree of sensitivity and convenience of the luciferase-mediated detection systems have created considerable interest in the development of firefly luciferase-based biosensors (see e.g., Green and Kricka,1984. Talanta 31: 173-176;

Blum, et al., 1989. J. Biolumin. Chemilumin: 4: 543-550).

[0053] The development of new reagents have made it possible to obtain stable light emission proportional to the concentrations of ATP (see e.g., Lundin, 1982. Applications of firefly luciferase In; Luminescent Assays (Raven Press, New York). With such stable light emission reagents, it is possible to make endpoint assays and to calibrate each individual assay by addition of a known amount of ATP. In addition, a stable light-emitting system also allows continuous monitoring of ATP-converting systems.

[0054] In a preferred embodiment, the ATP generating-ATP converting fusion protein is attached to an affinity tag. The term "affinity tag" is used herein to denote a peptide segment that can be attached to a polypeptide to provide for purification or detection of the polypeptide or provide sites for attachment of the polypeptide to a substrate. In principal, any peptide or protein for which an antibody or other specific binding agent is available can be used as an affinity tag. Affinity tags include a poly-histidine tract or a biotin carboxyl carrier protein (BCCP) domain, protein A (Nilsson et al., EMBO J. 4:1075, 1985; Nilsson et al., Methods Enzymol. 198:3, 1991), glutathione S transferase (Smith and Johnson, Gene 67:31, 1988), substance P, Flag.™. peptide (Hopp et al., Biotechnology 6:1204-1210, 1988; available from Eastman Kodak Co., New Haven, Conn.), streptavidin binding peptide, or other antigenic epitope or binding domain. See, in general Ford et al., Protein Expression and Purification 2: 95-107, 1991. DNAs encoding affinity tags are available from commercial suppliers (e.g., Pharmacia Biotech, Piscataway, N.J.).

[0055] As used herein, the term "poly-histidine tag," when used in reference to a fusion protein refers to the presence of two to ten histidine residues at either the amino- or carboxy-terminus of a protein of interest. A poly-histidine tract of six to ten residues is preferred. The poly-histidine tract is also defined functionally as being a number of consecutive histidine residues added to the protein of interest which allows the affinity purification of the resulting fusion protein on a nickel-chelate or IDA column.

[0056] In some embodiments, the fusion protein has an orientation such that the sulfurylase polypeptide is N-terminal to the luciferase polypeptide. In other embodiments, the luciferase polypeptide is N-terminal to the sulfurylase polypeptide. As used herein, the term sulfurylase-luciferase fusion protein refers to either of these orientations. The terms "amino-terminal" (N-terminal) and "carboxyl-terminal" (C-terminal) are used herein to denote positions within polypeptides and proteins. Where the context allows, these terms are used with reference to a particular sequence or portion of a polypeptide or protein to denote proximity or relative position. For example, a certain sequence positioned carboxyl-terminal to a reference sequence within a protein is located proximal to the carboxyl terminus of the reference sequence, but is not necessarily at the carboxyl terminus of the complete protein.

[0057] The fusion protein of this invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, e.g., by employing blunt-ended or "sticky"-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). The two polypeptides of the fusion protein can also be joined by a linker, such as a unique restriction site, which is engineered with specific primers during the cloning procedure. In one embodiment, the sulfurylase and luciferase polypeptides are joined by a linker, for example an ala-ala-ala linker which is encoded by a Not1 restriction site.

[0058] In one embodiment, the invention includes a recombinant polynucleotide that comprises a coding sequence for a fusion protein having an ATP generating polypeptide sequence and an ATP converting polypeptide sequence. In a preferred embodiment, the recombinant polynucleotide encodes a sulfurylase-luciferase fusion protein. The term "recombinant DNA molecule" or "recombinant polynucleotide" as used herein refers to a DNA molecule which is comprised of segments of DNA joined together by means of molecular biological techniques. The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule which is expressed from a recombinant DNA molecule.

[0059] In one aspect, this invention discloses a sulfurylase-luciferase fusion protein with an N-terminal hexahistidine tag and a BCCP tag. The nucleic acid sequence of the disclosed N-terminal hexahistidine-BCCP luciferase-sulfurylase gene (His6-BCCP L-S) gene is shown

His6-BCCP L-S Nucleotide Sequence (SEQ ID NO:3):

```
ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGGAAGCGCCAGCAGCA   60
GCGGAAATCAGTGGTCACATCGTACGTTCCCCGATGGTTGGTACTTTCTACCGCACCCCA  120
AGCCCGGACGCAAAAGCGTTCATCGAAGTGGGTCAGAAAGTCAACGTGGGCGATACCCTG  180
TGCATCGTTGAAGCCATGAAAATGATGAACCAGATCGAAGCGGACAAATCCGGTACCGTG  240
AAAGCAATTCTGGTCGAAAGTGGACAACCGGTAGAATTTGACGAGCCGCTGGTCGTCATC  300
GAGGGATCCGAGCTCGAGATCCAAATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCG  360
CCATTCTATCCTCTAGAGGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGA  420
TACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGAACATCACG  480
TACGCGGAATACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTG  540
AATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTG  600
TTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGAACGT  660
GAATTGCTCAACAGTATGAACATTTCGCAGCCTACCGTAGTGTTTGTTTCCAAAAAGGGG  720
TTGCAAAAAATTTTGAACGTGCAAAAAAAATTACCAATAATCCAGAAAATTATTATCATG  780
GATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTA  840
CCTCCCGGTTTTAATGAATACGATTTTGTACCAGAGTCCTTTGATCGTGACAAAACAATT  900
GCACTGATAATGAATTCCTCTGGATCTACTGGGTTACCTAAGGGTGTGGCCCTTCCGCAT  960
AGAACTGCCTGCGTCAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATT 1020
CCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACA 1080
CTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTG 1140
TTTTTACGATCCCTTCAGGATTACAAAATTCAAAGTGCGTTGCTAGTACCAACCCTATTT 1200
TCATTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAATT 1260
GCTTCTGGGGGCGCACCTCTTTCGAAAGAAGTCGGGGAAGCGGTTGCAAAACGCTTCCAT 1320
CTTCCAGGGATACGACAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACA 1380
CCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAG 1440
GTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAGAGAGGCGAATTATGTGTC 1500
AGAGGACCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATT 1560
GACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTC 1620
TTCATAGTTGACCGCTTGAAGTCTTTAATTAAATACAAAGGATATCAGGTGGCCCCCGCT 1680
GAATTGGAATCGATATTGTTACAACACCCCAACATCTTCGACGCGGGCGTGGCAGGTCTT 1740
CCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACG 1800
ATGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTG 1860
CGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCA 1920
AGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGTCCAAATTGGCGGCC 1980
GCTATGCCTGCTCCTCACGGTGGTATTCTACAAGACTTGATTGCTAGAGATGCGTTAAAG 2040
AAGAATGAATTGTTATCTGAAGCGCAATCTTCGGACATTTTAGTATGGAACTTGACTCCT 2100
AGACAACTATGTGATATTGAATTGATTCTAAATGGTGGGTTTTCTCCTCTGACTGGGTTT 2160
TTGAACGAAAACGATTACTCCTCTGTTGTTACAGATTCGAGATTAGCAGACGGCACATTG 2220
TGGACCATCCCTATTACATTAGATGTTGATGAAGCATTTGCTAACCAAATTAAACCAGAC 2280
ACAAGAATTGCCCTTTTCCAAGATGATGAAATTCCTATTGCTATACTTACTGTCCAGGAT 2340
```

GTTTACAAGCCAAACAAAACTATCGAAGCCGAAAAAGTCTTCAGAGGTGACCCAGAACAT 2400
CCAGCCATTAGCTATTTATTTAACGTTGCCGGTGATTATTACGTCGGCGGTTCTTTAGAA 2460
GCGATTCAATTACCTCAACATTATGACTATCCAGGTTTGCGTAAGACACCTGCCCAACTA 2520
AGACTTGAATTCCAATCAAGACAATGGGACCGTGTCGTAGCTTTCCAAACTCGTAATCCA 2580
ATGCATAGAGCCCACAGGGAGTTGACTGTGAGAGCCGCCAGAGAAGCTAATGCTAAGGTG 2640
CTGATCCATCCAGTTGTTGGACTAACCAAACCAGGTGATATAGACCATCACACTCGTGTT 2700
CGTGTCTACCAGGAAATTATTAAGCGTTATCCTAATGGTATTGCTTTCTTATCCCTGTTG 2760
CCATTAGCAATGAGAATGAGTGGTGATAGAGAAGCCGTATGGCATGCTATTATTAGAAAG 2820
AATTATGGTGCCTCCCACTTCATTGTTGGTAGAGACCATGCGGGCCCAGGTAAGAACTCC 2880
AAGGGTGTTGATTTCTACGGTCCATACGATGCTCAAGAATTGGTCGAATCCTACAAGCAT 2940
GAACTGGACATTGAAGTTGTTCCATTCAGAATGGTCACTTATTTGCCAGACGAAGACCGT 3000
TATGCTCCAATTGATCAAATTGACACCACAAAGACGAGAACCTTGAACATTTCAGGTACA 3060
GAGTTGAGACGCCGTTTAAGAGTTGGTGGTGAGATTCCTGAATGGTTCTCATATCCTGAA 3120
GTGGTTAAAATCCTAAGAGAATCCAACCCACCAAGACCAAAACAAGGTTTTTCAATTGTT 3180
TTAGGTAATTCATTAACCGTTTCTCGTGAGCAATTATCCATTGCTTTGTTGTCAACATTC 3240
TTGCAATTCGGTGGTGGCAGGTATTACAAGATCTTTGAACACAATAATAAGACAGAGTTA 3300
CTATCTTTGATTCAAGATTTCATTGGTTCTGGTAGTGGACTAATTATTCCAAATCAATGG 3360
GAAGATGACAAGGACTCTGTTGTTGGCAAGCAAAACGTTTACTTATTAGATACCTCAAGC 3420
TCAGCCGATATTCAGCTAGAGTCAGCGGATGAACCTATTTCACATATTGTACAAAAAGTT 3480
GTCCTATTCTTGGAAGACAATGGCTTTTTTGTATTTTAA 3519

[0060] The amino acid sequence of the disclosed His6-BCCP L-S polypeptide is presented using the three letter amino acid code (SEQ ID NO:4).

His6-BCCP L-S Amino Acid Sequence (SEQ ID NO:4)

Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Glu
1 5 10 15

Ala Pro Ala Ala Ala Glu Ile Ser Gly His Ile Val Arg Ser Pro Met
20 25 30

Val Gly Thr Phe Tyr Arg Thr Pro Ser Pro Asp Ala Lys Ala Phe Ile
35 40 45

Glu Val Gly Gln Lys Val Asn Val Gly Asp Thr Leu Cys Ile Val Glu
50 55 60

Ala Met Lys Met Met Asn Gln Ile Glu Ala Asp Lys Ser Gly Thr Val
65 70 75 80

Lys Ala Ile Leu Val Glu Ser Gly Gln Pro Val Glu Phe Asp Glu Pro
85 90 95

Leu Val Val Ile Glu Gly Ser Glu Leu Glu Ile Gln Met Glu Asp Ala
100 105 110

Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro Leu Glu Asp Gly

115 120 125

Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg Tyr Ala Leu Val
130 135 140

Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu Val Asn Ile Thr
145 150 155 160

Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala Glu Ala Met Lys
165 170 175

Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val Cys Ser Glu Asn
180 185 190

Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu Phe Ile Gly Val
195 200 205

Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg Glu Leu Leu Asn
210 215 220

Ser Met Asn Ile Ser Gln Pro Thr Val Val Phe Val Ser Lys Lys Gly
225 230 235 240

Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro Ile Ile Gln Lys
245 250 255

Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly Phe Gln Ser Met
260 265 270

Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe Asn Glu Tyr Asp
275 280 285

Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile Ala Leu Ile Met
290 295 300

Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val Ala Leu Pro His
305 310 315 320

Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp Pro Ile Phe Gly
325 330 335

Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val Val Pro Phe His
340 345 350

His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu Ile Cys Gly Phe
355 360 365

Arg Val Val Leu Met Tyr Arg Phe Glu Glu Glu Leu Phe Leu Arg Ser

370 375 380

Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val Pro Thr Leu Phe
385 390 395 400

Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr Asp Leu Ser Asn
405 410 415

Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser Lys Glu Val Gly
420 425 430

Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile Arg Gln Gly Tyr
435 440 445

Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr Pro Glu Gly Asp
450 455 460

Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe Phe Glu Ala Lys
465 470 475 480

Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val Asn Gln Arg Gly
485 490 495

Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly Tyr Val Asn Asn
500 505 510

Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly Trp Leu His Ser
515 520 525

Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe Phe Ile Val Asp
530 535 540

Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln Val Ala Pro Ala
545 550 555 560

Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile Phe Asp Ala Gly
565 570 575

Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu Pro Ala Ala Val
580 585 590

Val Val Leu Glu His Gly Lys Thr Met Thr Glu Lys Glu Ile Val Asp
595 600 605

Tyr Val Ala Ser Gln Val Thr Thr Ala Lys Lys Leu Arg Gly Gly Val
610 615 620

Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly Lys Leu Asp Ala

625 630 635 640

Arg Lys Ile Arg Glu Ile Leu Ile Lys Ala Lys Lys Gly Gly Lys Ser
645 650 655

Lys Leu Ala Ala Ala Met Pro Ala Pro His Gly Gly Ile Leu Gln Asp
660 665 670

Leu Ile Ala Arg Asp Ala Leu Lys Lys Asn Glu Leu Leu Ser Glu Ala
675 680 685

Gln Ser Ser Asp Ile Leu Val Trp Asn Leu Thr Pro Arg Gln Leu Cys
690 695 700

Asp Ile Glu Leu Ile Leu Asn Gly Gly Phe Ser Pro Leu Thr Gly Phe
705 710 715

Leu Asn Glu Asn Asp Tyr Ser Ser Val Val Thr Asp Ser Arg Leu Ala
720 725 730 735

Asp Gly Thr Leu Trp Thr Ile Pro Ile Thr Leu Asp Val Asp Glu Ala
740 745 750

Phe Ala Asn Gln Ile Lys Pro Asp Thr Arg Ile Ala Leu Phe Gln Asp
755 760 765

Asp Glu Ile Pro Ile Ala Ile Leu Thr Val Gln Asp Val Tyr Lys Pro
770 775 780

Asn Lys Thr Ile Glu Ala Glu Lys Val Phe Arg Gly Asp Pro Glu His
785 790 795

Pro Ala Ile Ser Tyr Leu Phe Asn Val Ala Gly Asp Tyr Tyr Val Gly
800 805 810 815

Gly Ser Leu Glu Ala Ile Gln Leu Pro Gln His Tyr Asp Tyr Pro Gly
820 825 830

Leu Arg Lys Thr Pro Ala Gln Leu Arg Leu Glu Phe Gln Ser Arg Gln
835 840 845

Trp Asp Arg Val Val Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala
850 855 860

His Arg Glu Leu Thr Val Arg Ala Ala Arg Glu Ala Asn Ala Lys Val
865 870 875

Leu Ile His Pro Val Val Gly Leu Thr Lys Pro Gly Asp Ile Asp His

880 885 890 895

His Thr Arg Val Arg Val Tyr Gln Glu Ile Ile Lys Arg Tyr Pro Asn
900 905 910

Gly Ile Ala Phe Leu Ser Leu Leu Pro Leu Ala Met Arg Met Ser Gly
915 920 925

Asp Arg Glu Ala Val Trp His Ala Ile Ile Arg Lys Asn Tyr Gly Ala
930 935 940

Ser His Phe Ile Val Gly Arg Asp His Ala Gly Pro Gly Lys Asn Ser
945 950 955

Lys Gly Val Asp Phe Tyr Gly Pro Tyr Asp Ala Gln Glu Leu Val Glu
960 965 970 975

Ser Tyr Lys His Glu Leu Asp Ile Glu Val Val Pro Phe Arg Met Val
980 985 990

Thr Tyr Leu Pro Asp Glu Asp Arg Tyr Ala Pro Ile Asp Gln Ile Asp
995 1000 1005

Thr Thr Lys Thr Arg Thr Leu Asn Ile Ser Gly Thr Glu Leu Arg Arg
1010 1015 1020

Arg Leu Arg Val Gly Gly Glu Ile Pro Glu Trp Phe Ser Tyr Pro Glu
1025 1030 1035

Val Val Lys Ile Leu Arg Glu Ser Asn Pro Pro Arg Pro Lys Gln Gly
1040 1045 1050 1055

Phe Ser Ile Val Leu Gly Asn Ser Leu Thr Val Ser Arg Glu Gln Leu
1060 1065 1070

Ser Ile Ala Leu Leu Ser Thr Phe Leu Gln Phe Gly Gly Gly Arg Tyr
1075 1080 1085

Tyr Lys Ile Phe Glu His Asn Asn Lys Thr Glu Leu Leu Ser Leu Ile
1090 1095 1100

Gln Asp Phe Ile Gly Ser Gly Ser Gly Leu Ile Ile Pro Asn Gln Trp
1105 1110 1115

Glu Asp Asp Lys Asp Ser Val Val Gly Lys Gln Asn Val Tyr Leu Leu
1120 1125 1130 1135

Asp Thr Ser Ser Ser Ala Asp Ile Gln Leu Glu Ser Ala Asp Glu Pro

1140 1145 1150

Ile Ser His Ile Val Gln Lys Val Val Leu Phe Leu Glu Asp Asn Gly
1155 1160 1165

Phe Phe Val Phe
1170

[0061] Accordingly, in one aspect, the invention provides for a fusion protein comprising a thermostable sulfurylase joined to at least one affinity tag. The nucleic acid sequence of the disclosed N-terminal hexahistidine-BCCP Bst ATP Sulfurylase (His6-BCCP Bst Sulfurylase) gene is shown below:

His6-BCCP Bst Sulfurylase Nucleotide Sequence (SEQ ID NO:5)

```
ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGGAAGCGCCAGCAGCA    60
GCGGAAATCAGTGGTCACATCGTACGTTCCCCGATGGTTGGTACTTTCTACCGCACCCCA   120
AGCCCGGACGCAAAAGCGTTCATCGAAGTGGGTCAGAAAGTCAACGTGGGCGATACCCTG   180
TGCATCGTTGAAGCCATGAAAATGATGAACCAGATCGAAGCGGACAAATCCGGTACCGTG   240
AAAGCAATTCTGGTCGAAAGTGGACAACCGGTAGAATTTGACGAGCCGCTGGTCGTCATC   300
GAGGGATCCGAGCTCGAGATCTGCAGCATGAGCGTAAGCATCCCGCATGGCGGCACATTG   360
ATCAACCGTTGGAATCCGGATTACCCAATCGATGAAGCAACGAAAACGATCGAGCTGTCC   420
AAAGCCGAACTAAGCGACCTTGAGCTGATCGGCACAGGCGCCTACAGCCCGCTCACCGGG   480
TTTTTAACGAAAGCCGATTACGATGCGGTCGTAGAAACGATGCGCCTCGCTGATGGCACT   540
GTCTGGAGCATTCCGATCACGCTGGCGGTGACGGAAGAAAAAGCGAGTGAACTCACTGTC   600
GGCGACAAAGCGAAACTCGTTTATGGCGGCGACGTCTACGGCGTCATTGAAATCGCCGAT   660
ATTTACCGCCCGGATAAAACGAAAGAAGCCAAGCTCGTCTATAAAACCGATGAACTCGCT   720
CACCCGGGCGTGCGCAAGCTGTTTGAAAAACCAGATGTGTACGTCGGCGGAGCGGTTACG   780
CTCGTCAAACGGACCGACAAAGGCCAGTTTGCTCCGTTTTATTTCGATCCGGCCGAAACG   840
CGGAAACGATTTGCCGAACTCGGCTGGAATACCGTCGTCGGCTTCCAAACACGCAACCCG   900
GTTCACCGCGCCCATGAATACATTCAAAAATGCGCGCTTGAAATCGTGGACGGCTTGTTT   960
TTAAACCCGCTCGTCGGCGAAACGAAAGCGGACGATATTCCGGCCGACATCCGGATGGAA  1020
AGCTATCAAGTGCTGCTGGAAAACTATTATCCGAAAGACCGCGTTTTCTTGGGCGTCTTC  1080
CAAGCTGCGATGCGCTATGCCGGTCCGCGCGAAGCGATTTTCCATGCCATGGTGCGGAAA  1140
AACTTCGGCTGCACGCACTTCATCGTCGGCCGCGACCATGCGGGCGTCGGCAACTATTAC  1200
GGCACGTATGATGCGCAAAAAATCTTCTCGAACTTTACAGCCGAAGAGCTTGGCATTACA  1260
CCGCTCTTTTTCGAACACAGCTTTTATTGCACGAAATGCGAAGGCATGGCATCGACGAAA  1320
ACATGCCCGCACGACGCACAATATCACGTTGTCCTTTCTGGCACGAAAGTCCGTGAAATG  1380
TTGCGTAACGGCCAAGTGCCGCCGAGCACATTCAGCCGTCCGGAAGTGGCCGCCGTTTTG  1440
ATCAAAGGGCTGCAAGAACGCGAAACGGTCGCCCCGTCAGCGCGCTAA              1488
```

[0062] The amino acid sequence of the His6-BCCP Bst Sulfurylase polypeptide is presented using the three letter amino acid code in Table 6 (SEQ ID NO:6).

His6-BCCP Bst Sulfurylase Amino Acid Sequence (SEQ ID NO:6)

Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Glu
1 5 10 15

Ala Pro Ala Ala Ala Glu Ile Ser Gly His Ile Val Arg Ser Pro Met
20 25 30

Val Gly Thr Phe Tyr Arg Thr Pro Ser Pro Asp Ala Lys Ala Phe Ile
35 40 45

Glu Val Gly Gln Lys Val Asn Val Gly Asp Thr Leu Cys Ile Val Glu
50 55 60

Ala Met Lys Met Met Asn Gln Ile Glu Ala Asp Lys Ser Gly Thr Val
65 70 75 80

Lys Ala Ile Leu Val Glu Ser Gly Gln Pro Val Glu Phe Asp Glu Pro
85 90 95

Leu Val Val Ile Glu Gly Ser Glu Leu Glu Ile Cys Ser Met Ser Val
100 105 110

Ser Ile Pro His Gly Gly Thr Leu Ile Asn Arg Trp Asn Pro Asp Tyr
115 120 125

Pro Ile Asp Glu Ala Thr Lys Thr Ile Glu Leu Ser Lys Ala Glu Leu
130 135 140

Ser Asp Leu Glu Leu Ile Gly Thr Gly Ala Tyr Ser Pro Leu Thr Gly
145 150 155 160

Phe Leu Thr Lys Ala Asp Tyr Asp Ala Val Val Glu Thr Met Arg Leu
165 170 175

Ala Asp Gly Thr Val Trp Ser Ile Pro Ile Thr Leu Ala Val Thr Glu
180 185 190

Glu Lys Ala Ser Glu Leu Thr Val Gly Asp Lys Ala Lys Leu Val Tyr
195 200 205

Gly Gly Asp Val Tyr Gly Val Ile Glu Ile Ala Asp Ile Tyr Arg Pro
210 215 220

Asp Lys Thr Lys Glu Ala Lys Leu Val Tyr Lys Thr Asp Glu Leu Ala
225 230 235 240

His Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val Tyr Val Gly
245 250 255

Gly Ala Val Thr Leu Val Lys Arg Thr Asp Lys Gly Gln Phe Ala Pro
260 265 270

Phe Tyr Phe Asp Pro Ala Glu Thr Arg Lys Arg Phe Ala Glu Leu Gly
275 280 285

Trp Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val His Arg Ala
290 295 300

His Glu Tyr Ile Gln Lys Cys Ala Leu Glu Ile Val Asp Gly Leu Phe
305 310 315 320

Leu Asn Pro Leu Val Gly Glu Thr Lys Ala Asp Asp Ile Pro Ala Asp
325 330 335

Ile Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr Tyr Pro Lys
340 345 350

Asp Arg Val Phe Leu Gly Val Phe Gln Ala Ala Met Arg Tyr Ala Gly
355 360 365

Pro Arg Glu Ala Ile Phe His Ala Met Val Arg Lys Asn Phe Gly Cys
370 375 380

Thr His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly Asn Tyr Tyr
385 390 395 400

Gly Thr Tyr Asp Ala Gln Lys Ile Phe Ser Asn Phe Thr Ala Glu Glu
405 410 415

Leu Gly Ile Thr Pro Leu Phe Phe Glu His Ser Phe Tyr Cys Thr Lys
420 425 430

Cys Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp Ala Gln Tyr
435 440 445

His Val Val Leu Ser Gly Thr Lys Val Arg Glu Met Leu Arg Asn Gly
450 455 460

Gln Val Pro Pro Ser Thr Phe Ser Arg Pro Glu Val Ala Ala Val Leu
465 470 475 480

Ile Lys Gly Leu Gln Glu Arg Glu Thr Val Ala Pro Ser Ala Arg
485 490 495

[0063] Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an ATP generating polypeptide and an ATP converting polypeptide, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are

capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors *(e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

[0064] The recombinant expression vectors of the invention comprise a nucleic acid in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce a fusion protein.

[0065] The recombinant expression vectors of the invention can be designed for expression of the fusion protein in prokaryotic or eukaryotic cells. For example, a sulfurylase-luciferase fusion protein can be expressed in bacterial cells such as *E. coli,* insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro,* for example using T7 promoter regulatory sequences and T7 polymerase.

[0066] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein.

[0067] In another embodiment, the ATP generating-ATP converting fusion protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, et al., (1987) EMBO J 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al., (1987) Gene 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

[0068] Alternatively, the fusion protein can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells *(e.g.,* SF9 cells) include the pAc series (Smith et al. (1983) Mol Cell Biol 3:2156-2165) and the pVL series (Lucklow and Summers (1989) Virology 170:31-39).

[0069] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) Nature 329:840) and pMT2PC (Kaufman et al. (1987) EMBO J 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells. See, *e.g.,* Chapters 16 and 17 of Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0070] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type *(e.g.,* tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) Genes Dev 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) Adv Immunol 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) EMBO J 8:729-733) and immunoglobulins (Banerji et al. (1983) Cell 33:729-740; Queen and Baltimore (1983) Cell 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) PNAS 86:5473-5477), pancreas-specific promoters *(*Edlund et al. (1985) Science 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss (1990) Science 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) Genes Dev 3:537-546).

[0071] Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. The invention also includes a kit comprising a sulfurylase-luciferase fusion protein expression vector.

[0072] A host cell can be any prokaryotic or eukaryotic cell. For example, the sulfurylase-luciferase fusion protein can be expressed in bacterial cells such as *E. coli,* insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0073] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

[0074] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding ORFX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0075] A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) the fusion protein. Accordingly, the invention further provides methods for producing the fusion protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding the fusion protein has been introduced) in a suitable medium such that the fusion protein is produced. In another embodiment, the method further comprises isolating the fusion protein from the medium or the host cell.

[0076] The invention also includes a fusion protein bound to a mobile support. In a preferred embodiment, the fusion gene is a sulfurylase-luciferase fusion gene. In another embodiment, the mobile support is bound to strepavidin. The mobile support could be a bead or optical fiber. In a preferred embodiment, the bead is a nickel-agarose bead or a MPG-Streptavidin bead. In one embodiment, the sulfurylase-luciferase fusion protein is bound to the beads in a 1:3 ratio of protein to bead. It can be attached to the solid support via a covalent or non-covalent interaction. In general, any linkage recognized in the art can be used. Examples of such linkages common in the art include any suitable metal (*eg.*, Co^{2+}, Ni^{2+})-hexahistidine complex, a biotin binding protein, *e.g.*, NEUTRAVIDIN™ modified avidin (Pierce Chemicals, Rockford, IL), streptavidin/biotin, avidin/biotin, glutathione S-transferase (GST)/glutathione, monoclonal antibody/antigen, and maltose binding protein/maltose, and pluronic coupling technologies. Samples containing the appropriate tag are incubated with the sensitized substrate so that zero, one, or multiple molecules attach at each sensitized site.

[0077] Acetyl-CoA carboxylase (ACCase) catalyzes the first committed step in de novo fatty acid biosynthesis. It belongs to a group of carboxylases that use biotin as cofactor and bicarbonate as a source of the carboxyl group. There are two types of ACCase: prokaryotic ACCase (e.g., E. *coli, P. aeruginosa, Anabaena, Synechococcus* and probably pea chloroplast) in which the three functional domains: biotin carboxylase (BC), biotin carboxyl carrier protein (BCCP) and carboxyltransferase (CT) are located on separable subunits and eukaryotic ACCase (e.g., rat, chicken, yeast, diatom and wheat) in which all the domains are located on one large polypeptide. It is known that a BCCP as a subunit of acetyl CoA carboxylase from E. *coli* is biotinated at the Lys residue at the 122-position by the action of biotin holoenzyme synthetase in E. *coli* (Journal of Biological Chemistry, 263, 6461 (1988)). In a preferred embodiment of this invention, the fusion protein is bound to a BCCP domain which is then utilized for binding avidins; therefore, it can bind to a streptavidin mobile support. One biotin-(strept-)avidin-based anchoring method uses a thin layer of a photoactivatable biotin analog dried onto a solid surface. (Hengsakul and Cass, 1996. Bioconjugate Chem. 7: 249-254). The biotin analog is then exposed to white light through a mask, so as to create defined areas of activated biotin. Avidin (or streptavidin) is then added and allowed to bind to the activated biotin. The avidin possesses free biotin binding sites which can be utilized to "anchor" the biotinylated proteins through a biotin-(strept-)avidin linkage.

[0078] Alternatively, the fusion protein can be attached to the solid support with a biotin derivative possessing a photoremovable protecting group. This moiety is covalently bound to bovine serum albumin (BSA), which is attached to the solid support, *e.g.*, a glass surface. See Pirrung and Huang, 1996. Bioconjugate Chem. 7: 317-321. A mask is then used to create activated biotin within the defined irradiated areas. Avidin may then be localized to the irradiated area, with a

biotinylated sulfurylase-luciferase fusion protein subsequently attached through a BSA-biotin-avidin-biotin link.

[0079] Another method of attachment is with the use of a pluronics based attachment. Pluronics attach to hydrophobic surfaces by virtue of the reaction between the hydrophobic surface and the polypropylene oxide. The remaining polyethylene oxide groups extend off the surface, thereby creating a hydrophilic environment. Nitrilotriacetic acid (NTA) can be conjugated to the terminal ends of the polyethylene oxide chains to allow for hexahistidine tagged proteins to be attached.

[0080] This invention provides methods of sequencing which utilize and ATP generating polypeptide-ATP converting polypeptide fusion protein for detection. In a preferred embodiment, the nucleotide sequence of the sequencing product is determined by measuring inorganic pyrophosphate (PPi) liberated from a nucleotide triphosphate (dNTP) as the dNMP is incorporated into an extended sequence primer. This method of sequencing is termed Pyrosequencing™ technology (PyroSequencing AB, Stockholm, Sweden). It can be performed in solution (liquid phase) or as a solid phase technique. Various sequencing methods, including PPi sequencing methods, are described in, e.g., W09813523A1, Ronaghi, et al., 1996. Anal. Biochem. 242: 84-89, and Ronaghi, et al., 1998. Science 281: 363-365 (1998), US patent 6,274,320 and the patent application USSN 10/104,280 which was filed on March 21, 2001 (21465-501CIP3). These disclosures of sequencing are incorporated herein in their entirety, by reference.

[0081] Pyrophosphate released under these conditions can be detected enzymatically (e.g., by the generation of light in the luciferase-luciferin reaction). Such methods enable a nucleotide to be identified in a given target position, and the DNA to be sequenced simply and rapidly while avoiding the need for electrophoresis and the use of potentially dangerous radiolabels.

[0082] The invention also provides a method for sequencing nucleic acids which generally comprises (a) providing one or more nucleic acid anchor primers and a plurality of single-stranded circular nucleic acid templates disposed within a plurality of reaction chambers or cavities; (b) annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (c) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (d) annealing an effective amount of a sequencing primer to one or more copies of said covalently linked complementary nucleic acid; (e) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and (f) identifying the PPi sequencing reaction byproduct with the use of an ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the sequence of the nucleic acid. In one embodiment, a dATP or ddATP analogue is used in place of deoxy- or dideoxy adenosine triphosphate. This analogue is capable of acting as a substrate for a polymerase but incapable of acting as a substrate for a PPi-detection enzyme. This method can be carried out in separate parallel common reactions in an aqueous environment.

[0083] In another aspect, the invention includes a method of determining the base sequence of a plurality of nucleotides on an array, which generally comprises (a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface; (b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (c) utilizing an ATP generating polypeptide-ATP converting polypeptide fusion protein to detect whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands in which incorporation of the nucleoside 5'-triphosphate precursor indicates that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and (d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and (e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

[0084] The anchor primers of the invention generally comprise a stalk region and at least one adaptor region. In a preferred embodiment the anchor primer contains at least two contiguous adapter regions. The stalk region is present at the 5' end of the anchor primer and includes a region of nucleotides for attaching the anchor primer to the solid substrate.

[0085] The adaptor region(s) comprise nucleotide sequences that hybridize to a complementary sequence present in one or more members of a population of nucleic acid sequences. In some embodiments, the anchor primer includes two adjoining adaptor regions, which hybridize to complementary regions ligated to separate ends of a target nucleic acid sequence. In additional embodiments, the adapter regions in the anchor primers are complementary to non-contiguous regions of sequence present in a second nucleic acid sequence. Each adapter region, for example, can be

homologous to each terminus of a fragment produced by digestion with one or more restriction endonucleases. The fragment can include, e.g., a sequence known or suspected to contain a sequence polymorphism. Additionally, the anchor primer may contain two adapter regions that are homologous to a gapped region of a target nucleic acid sequence, i. e., one that is non-contiguous because of a deletion of one or more nucleotides. When adapter regions having these sequences are used, an aligning oligonucleotide corresponding to the gapped sequence may be annealed to the anchor primer along with a population of template nucleic acid molecules.

[0086] The anchor primer may optionally contain additional elements such as one or more restriction enzyme recognition sites, RNA polymerase binding sites, e.g., a T7 promoter site, or sequences present in identified DNA sequences, e.g., sequences present in known genes. The adapter region(s) may also include sequences known to flank sequence polymorphisms. Sequence polymorphisms include nucleotide substitutions, insertions, deletions, or other rearrangements which result in a sequence difference between two otherwise identical nucleic acid sequences. An example of a sequence polymorphism is a single nucleotide polymorphism (SNP).

[0087] In general, any nucleic acid capable of base-pairing can be used as an anchor primer. In some embodiments, the anchor primer is an oligonucleotide. As utilized herein the term oligonucleotide includes linear oligomers of natural or modified monomers or linkages, e.g., deoxyribonucleosides, ribonucleosides, anomeric forms thereof, peptide nucleic acids (PNAs), and the like, that are capable of specifically binding to a target polynucleotide by way of a regular pattern of monomer-to-monomer interactions. These types of interactions can include, e.g., Watson-Crick type of base-pairing, base stacking, Hoogsteen or reverse-Hoogsteen types of base-pairing, or the like. Generally, the monomers are linked by phosphodiester bonds, or analogs thereof, to form oligonucleotides ranging in size from, *e.g.*, 3-200, 8-150, 10-100, 20-80, or 25-50 monomeric units. Whenever an oligonucleotide is represented by a sequence of letters, it is understood that the nucleotides are oriented in the 5' $\rightarrow$' direction, from left-to-right, and that the letter "A" donates deoxyadenosine, the letter "T" denotes thymidine, the letter "C" denotes deoxycytosine, and the letter "G" denotes deoxyguanosine, unless otherwise noted herein. The oligonucleotides of the present invention can include non-natural nucleotide analogs. However, where, for example, processing by enzymes is required, or the like, oligonucleotides comprising naturally occurring nucleotides are generally required for maintenance of biological function.

[0088] Anchor primers are linked to the solid substrate at the sensitized sites. They can be linked by the same method of linkage as described for the fusion protein to the solid support. A region of a solid substrate containing a linked primer is referred to herein as an anchor pad. Thus, by specifying the sensitized states on the solid support, it is possible to form an array or matrix of anchor pads. The anchor pads can be, e.g., small diameter spots etched at evenly spaced intervals on the solid support. The anchor pads can be located at the bottoms of the cavitations or wells if the substrate has been cavitated, etched, or otherwise micromachined as discussed above.

[0089] In one embodiment, the anchor primer is linked to a particle. The anchor primer can be linked to the particle prior to formation of the extended anchor primer or after formation of the extended anchor primer.

[0090] Each sensitized site on a solid support is potentially capable of attaching multiple anchor primers. Thus, each anchor pad may include one or more anchor primers. It is preferable to maximize the number of pads that have only a single productive reaction center (e.g., the number of pads that, after the extension reaction, have only a single sequence extended from the anchor primer). This can be accomplished by techniques which include, but are not limited to: (i) varying the dilution of biotinylated anchor primers that are washed over the surface; (*ii*) varying the incubation time that the biotinylated primers are in contact with the avidin surface; (iii) varying the concentration of open- or closed-circular template so that, on average, only one primer on each pad is extended to generate the sequencing template; or (*iv*) reducing the size of the anchor pad to approach single-molecule dimensions (< 1 μm) such that binding of one anchor inhibits or blocks the binding of another anchor (*e.g.* by photoactivation of a small spot); or (v) reducing the size of the anchor pad such that binding of one circular template inhibits or blocks the binding of a second circular template.

[0091] In some embodiments, each individual pad contains just one linked anchor primer. Pads having only one anchor primer can be made by performing limiting dilutions of a selected anchor primer on to the solid support such that, on average, only one anchor primer is deposited on each pad. The concentration of anchor primer to be applied to a pad can be calculated utilizing, for example, a Poisson distribution model.

[0092] In order to maximize the number of reaction pads that contain a single anchor primer, a series of dilution experiments are performed in which a range of anchor primer concentrations or circular template concentrations are varied. For highly dilute concentrations of primers, primers and circular templates binding to the same pad will be independent of each other, and a Poisson distribution will characterize the number of anchor primers extended on any one pad. Although there will be variability in the number of primers that are actually extended, a maximum of 37% of the pads will have a single extended anchor primer (the number of pads with a single anchor oligonucleotide).

[0093] In other embodiments multiple anchor primers are attached to any one individual pad in an array. Limiting dilutions of a plurality of circular nucleic acid templates (described in more detail below) may be hybridized to the anchor primers so immobilized such that, on average, only one primer on each pad is hybridized to a nucleic acid template. Library concentrations to be used may be calculated utilizing, for example, limiting dilutions and a Poisson distribution model.

[0094] The nucleic acid templates that can be sequenced according to the invention, e.g., a nucleic acid library, in general can include open circular or closed circular nucleic acid molecules. A "closed circle" is a covalently closed circular nucleic acid molecule, e.g., a circular DNA or RNA molecule. An "open circle" is a linear single-stranded nucleic acid molecule having a 5' phosphate group and a 3' hydroxyl group. In one embodiment, the single stranded nucleic acid contains at least 100 copies of nucleic acid sequence, each copy covalently linked end to end. In some embodiments, the open circle is formed *in situ* from a linear double-stranded nucleic acid molecule. The ends of a given open circle nucleic acid molecule can be ligated by DNA ligase. Sequences at the 5' and 3' ends of the open circle molecule are complementary to two regions of adjacent nucleotides in a second nucleic acid molecule, *e.g.*, an adapter region of an anchor primer, or to two regions that are nearly adjoining in a second DNA molecule. Thus, the ends of the open-circle molecule can be ligated using DNA ligase, or extended by DNA polymerase in a gap-filling reaction. Open circles are described in detail in Lizardi, U.S. Pat. No. 5,854,033. An open circle can be converted to a closed circle in the presence of a DNA ligase (for DNA) or RNA ligase following, *e.g.*, annealing of the open circle to an anchor primer.

[0095] If desired, nucleic acid templates can be provided as padlock probes. Padlock probes are linear oligonucleotides that include target-complementary sequences located at each end, and which are separated by a linker sequence. The linkers can be ligated to ends of members of a library of nucleic acid sequences that have been, *e.g.*, physically sheared or digested with restriction endonucleases. Upon hybridization to a target-sequence, the 5'- and 3'-terminal regions of these linear oligonucleotides are brought in juxtaposition. This juxtaposition allows the two probe segments (if properly hybridized) to be covalently-bound by enzymatic ligation (*e.g.*, with T4 DNA ligase), thus converting the probes to circularly-closed molecules which are catenated to the specific target sequences (see *e.g.*, Nilsson, et al., 1994. Science 265: 2085-2088). The resulting probes are suitable for the simultaneous analysis of many gene sequences both due to their specificity and selectivity for gene sequence variants (see e.g., Lizardi, et al., 1998. Nat. Genet. 19: 225-232; Nilsson, et al., 1997. Nat. Genet. 16: 252-255) and due to the fact that the resulting reaction products remain localized to the specific target sequences. Moreover, intramolecular ligation of many different probes is expected to be less susceptible to non-specific cross-reactivity than multiplex PCR-based methodologies where non-cognate pairs of primers can give rise to irrelevant amplification products (see e.g., Landegren and Nilsson, 1997. Ann. Med. 29: 585-590).

[0096] A starting library can be constructed comprising either single-stranded or double-stranded nucleic acid molecules, provided that the nucleic acid sequence includes a region that, if present in the library, is available for annealing, or can be made available for annealing, to an anchor primer sequence. For example, when used as a template for rolling circle amplification, a region of a double-stranded template needs to be at least transiently single-stranded in order to act as a template for extension of the anchor primer.

[0097] Library templates can include multiple elements, including, but not limited to, one or more regions that are complementary to the anchor primer. For example, the template libraries may include a region complementary to a sequencing primer, a control nucleotide region, and an insert sequence comprised of the sequencing template to be subsequently characterized. As is explained in more detail below, the control nucleotide region is used to calibrate the relationship between the amount of byproduct and the number of nucleotides incorporated. As utilized herein the term "complement" refers to nucleotide sequences that are able to hybridize to a specific nucleotide sequence to form a matched duplex.

[0098] In one embodiment, a library template includes: (*i*) two distinct regions that are complementary to the anchor primer, (*ii*) one region homologous to the sequencing primer, (*iii*) one optional control nucleotide region, (*iv*) an insert sequence of, *e.g.,* 30-500, 50-200, or 60-100 nucleotides, that is to be sequenced. The template can, of course, include two, three, or all four of these features.

[0099] The template nucleic acid can be constructed from any source of nucleic acid, e.g., any cell, tissue, or organism, and can be generated by any art-recognized method. Suitable methods include, e.g., sonication of genomic DNA and digestion with one or more restriction endonucleases (RE) to generate fragments of a desired range of lengths from an initial population of nucleic acid molecules. Preferably, one or more of the restriction enzymes have distinct four-base recognition sequences. Examples of such enzymes include, e.g., Sau3A1, MspI, and TaqI. Preferably, the enzymes are used in conjunction with anchor primers having regions containing recognition sequences for the corresponding restriction enzymes. In some embodiments, one or both of the adapter regions of the anchor primers contain additional sequences adjoining known restriction enzyme recognition sequences, thereby allowing for capture or annealing to the anchor primer of specific restriction fragments of interest to the anchor primer. In other embodiments, the restriction enzyme is used with a type IIS restriction enzyme.

[0100] Alternatively, template libraries can be made by generating a complementary DNA (cDNA) library from RNA, e.g., messenger RNA (mRNA). The cDNA library can, if desired, be further processed with restriction endonucleases to obtain a 3' end characteristic of a specific RNA, internal fragments, or fragments including the 3' end of the isolated RNA. Adapter regions in the anchor primer may be complementary to a sequence of interest that is thought to occur in the template library, e.g., a known or suspected sequence polymorphism within a fragment generated by endonuclease digestion.

[0101] In one embodiment, an indexing oligonucleotide can be attached to members of a template library to allow for

subsequent correlation of a template nucleic acid with a population of nucleic acids from which the template nucleic acid is derived. For example, one or more samples of a starting DNA population can be fragmented separately using any of the previously disclosed methods (e.g., restriction digestion, sonication). An indexing oligonucleotide sequence specific for each sample is attached to, e.g., ligated to, the termini of members of the fragmented population. The indexing oligonucleotide can act as a region for circularization, amplification and, optionally, sequencing, which permits it to be used to index, or code, a nucleic acid so as to identify the starting sample from which it is derived.

[0102] Distinct template libraries made with a plurality of distinguishable indexing primers can be mixed together for subsequent reactions. Determining the sequence of the member of the library allows for the identification of a sequence corresponding to the indexing oligonucleotide. Based on this information, the origin of any given fragment can be inferred.

[0103] Libraries of nucleic acids are annealed to anchor primer sequences using recognized techniques (see, e.g., Hatch, et al., 1999. Genet. Anal. Biomol. Engineer. 15: 35-40; Kool, U.S. Patent No. 5,714, 320 and Lizardi, U.S. Patent No. 5,854,033). In general, any procedure for annealing the anchor primers to the template nucleic acid sequences is suitable as long as it results in formation of specific, i. e., perfect or nearly perfect, complementarity between the adapter region or regions in the anchor primer sequence and a sequence present in the template library.

[0104] A number of *in vitro* nucleic acid amplification techniques may be utilized to extend the anchor primer sequence. The size of the amplified DNA preferably is smaller than the size of the anchor pad and also smaller than the distance between anchor pads.

[0105] The amplification is typically performed in the presence of a polymerase, e.g., a DNA or RNA-directed DNA polymerase, and one, two, three, or four types of nucleotide triphosphates, and, optionally, auxiliary binding proteins. In general, any polymerase capable of extending a primed 3'-OH group can be used a long as it lacks a 3' to 5' exonuclease activity. Suitable polymerases include, *e.g.*, the DNA polymerases from *Bacillus stearothermophilus, Thermus acquaticus, Pyrococcus furiosis, Thermococcus litoralis,* and *Thermus thermophilus,* bacteriophage T4 and T7, and the *E. coli* DNA polymerase I Klenow fragment. Suitable RNA-directed DNA polymerases include, *e.g.,* the reverse transcriptase from the Avian Myeloblastosis Virus, the reverse transcriptase from the Moloney Murine Leukemia Virus, and the reverse transcriptase from the Human Immunodeficiency Virus-I.

[0106] A number of *in vitro* nucleic acid amplification techniques have been described. These amplification methodologies may be differentiated into those methods: (i) which require temperature cycling - polymerase chain reaction (PCR) (see *e.g.,* Saiki, et al., 1995. Science 230: 1350-1354), ligase chain reaction (see *e.g.,* Barany, 1991. Proc. Natl. Acad. Sci. USA 88: 189-193; Barringer, et al., 1990. Gene 89: 117-122) and transcription-based amplification (see e.g., Kwoh, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 1173-1177) and (*ii*) isothermal amplification systems - self-sustaining, sequence replication (see e.g., Guatelli, et al., 1990. Proc. Natl. Acad. Sci. USA 87: 1874-1878); the Qβ replicase system (see *e.g.,* Lizardi, et al., 1988. BioTechnology 6: 1197-1202); strand displacement amplification Nucleic Acids Res. 1992 Apr 11;20(7):1691-6.; and the methods described in PNAS 1992 Jan 1;89(1):392-6; and NASBA J Virol Methods. 1991 Dec;35(3):273-86.

[0107] Isothermal amplification also includes rolling circle-based amplification (RCA). RCA is discussed in, *e.g.,* Kool, U.S. Patent No. 5,714,320 and Lizardi, U.S. Patent No. 5,854,033; Hatch, et al., 1999. Genet. Anal. Biomol. Engineer. 15: 35-40. The result of the RCA is a single DNA strand extended from the 3' terminus of the anchor primer (and thus is linked to the solid support matrix) and including a concatamer containing multiple copies of the circular template annealed to a primer sequence. Typically, 1,000 to 10,000 or more copies of circular templates, each having a size of, e.g., approximately 30-500, 50-200, or 60-100 nucleotides size range, can be obtained with RCA.

[0108] *In vivo,* RCR is utilized in several biological systems. For example, the genome of several bacteriophage are single-stranded, circular DNA. During replication, the circular DNA is initially converted to a duplex form, which is then replicated by the aforementioned rolling-circle replication mechanism. The displaced terminus generates a series of genomic units that can be cleaved and inserted into the phage particles. Additionally, the displaced single-strand of a rolling-circle can be converted to duplex DNA by synthesis of a complementary DNA strand. This synthesis can be used to generate the concatemeric duplex molecules required for the maturation of certain phage DNAs. For example, this provides the principle pathway by which λ, bacteriophage matures. RCR is also used *in vivo* to generate amplified rDNA in *Xenopus* oocytes, and this fact may help explain why the amplified rDNA is comprised of a large number of identical repeating units. In this case, a single genomic repeating unit is converted into a rolling-circle. The displaced terminus is then converted into duplex DNA which is subsequently cleaved from the circle so that the two termini can be ligated together so as to generate the amplified circle of rDNA.

[0109] Through the use of the RCA reaction, a strand may be generated which represents many tandem copies of the complement to the circularized molecule. For example, RCA has recently been utilized to obtain an isothermal cascade amplification reaction of circularized padlock probes *in vitro* in order to detect single-copy genes in human genomic DNA samples (see Lizardi, et al., 1998. Nat. Genet. 19: 225-232). In addition, RCA has also been utilized to detect single DNA molecules in a solid phase-based assay, although difficulties arose when this technique was applied to *in situ* hybridization (see Lizardi, et al., 1998. Nat. Genet. 19: 225-232).

[0110] If desired, RCA can be performed at elevated temperatures, e.g., at temperatures greater than 37˚ C, 42˚ C,

45˚ C, 50˚ C, 60˚ C, or 70˚ C. In addition, RCA can be performed initially at a lower temperature, e.g., room temperature, and then shifted to an elevated temperature. Elevated temperature RCA is preferably performed with thermostable nucleic acid polymerases and with primers that can anneal stably and with specificity at elevated temperatures.

RCA can also be performed with non-naturally occurring oligonucleotides, e.g., peptide nucleic acids. Further, RCA can be performed in the presence of auxiliary proteins such as single-stranded binding proteins.

[0111] The development of a method of amplifying short DNA molecules which have been immobilized to a solid support, termed RCA has been recently described in the literature (see *e.g.,* Hatch, et al., 1999. Genet. Anal. Biomol. Engineer. 15: 35-40; Zhang, et al., 1998. Gene 211: 277-85; Baner, et al., 1998. Nucl. Acids Res. 26: 5073-5078; Liu, et al., 1995. J. Am. Chem. Soc. 118: 1587-1594; Fire and Xu, 1995. Proc. Natl. Acad. Sci. USA 92: 4641-4645; Nilsson, et al., 1994. Science 265: 2085-2088). RCA targets specific DNA sequences through hybridization and a DNA ligase reaction. The circular product is then subsequently used as a template in a rolling circle replication reaction.

[0112] RCA driven by DNA polymerase can replicate circularized oligonucleotide probes with either linear or geometric kinetics under isothermal conditions. In the presence of two primers (one hybridizing to the + strand, and the other, to the - strand of DNA), a complex pattern of DNA strand displacement ensues which possesses the ability to generate $1x10^9$ or more copies of each circle in a short period of time (*i.e.,* less-than 90 minutes), enabling the detection of single-point mutations within the human genome. Using a single primer, RCA generates hundreds of randomly-linked copies of a covalently closed circle in several minutes. If solid support matrix-associated, the DNA product remains bound at the site of synthesis, where it may be labeled, condensed, and imaged as a point light source. For example, linear oligonucleotide probes, which can generate RCA signals, have been bound covalently onto a glass surface. The color of the signal generated by these probes indicates the allele status of the target, depending upon the outcome of specific, target-directed ligation events. As RCA permits millions of individual probe molecules to be counted and sorted, it is particularly amenable for the analysis of rare somatic mutations. RCA also shows promise for the detection of padlock probes bound to single-copy genes in cytological preparations.

[0113] In addition, a solid-phase RCA methodology has also been developed to provide an effective method of detecting constituents within a solution. Initially, a recognition step is used to generate a complex h a circular template is bound to a surface. A polymerase enzyme is then used to amplify the bound complex. RCA uses small DNA probes that are amplified to provide an intense signal using detection methods, including the methods described in more detail below. Other examples of isothermal amplification systems include, e.g., (i) self-sustaining, sequence replication (see e.g., Guatelli, et al., 1990. Proc. Natl. Acad. Sci. USA 87: 1874-1878), (*ii*) the Qβ replicase system (see *e.g.,* Lizardi, et al., 1988. BioTechnology 6: 1197-1202), and (*iii*) nucleic acid sequence-based amplification (NASBA™; see Kievits, et al., 1991. J. Virol. Methods 35: 273-286).

[0114] Amplification of a nucleic acid template as described above results in multiple copies of a template nucleic acid sequence covalently linked to an anchor primer. In one embodiment, a region of the sequence product is determined by annealing a sequencing primer to a region of the template nucleic acid, and then contacting the sequencing primer with a DNA polymerase and a known nucleotide triphosphate, *i.e.,* dATP, dCTP, dGTP, dTTP, or an analog of one of these nucleotides. The sequence can be determined by detecting a sequence reaction byproduct, as is described below.

[0115] The sequence primer can be any length or base composition, as long as it is capable of specifically annealing to a region of the amplified nucleic acid template. No particular structure for the sequencing primer is required so long as it is able to specifically prime a region on the amplified template nucleic acid. Preferably, the sequencing primer is complementary to a region of the template that is between the sequence to be characterized and the sequence hybridizable to the anchor primer. The sequencing primer is extended with the DNA polymerase to form a sequence product. The extension is performed in the presence of one or more types of nucleotide triphosphates, and if desired, auxiliary binding proteins.

[0116] The method comprises the steps of (a) introducing the template nucleic acid polymer into a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added; (b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released; (c) separately recovering each of the feedstocks from the polymerization environment; and (d) measuring the amount of inorganic pyrophosphate by utilizing an ATP generating polypeptide-ATP converting polypeptide fusion protein in each of the recovered feedstocks to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer.

[0117] The sequence primer can be any length or base composition, as long as it is capable of specifically annealing to a region of the amplified nucleic acid template. No particular structure is required for the sequencing primer so long as it is able to specifically prime a region on the amplified template nucleic acid. Preferably, the sequencing primer is complementary to a region of the template that is between the sequence to be characterized and the sequence hybridizable to the anchor primer. The sequencing primer is extended with the DNA polymerase to form a sequence product.

The extension is performed in the presence of one or more types of nucleotide triphosphates, and if desired, auxiliary binding proteins.

[0118] This invention also includes a method wherein the amount of inorganic pyrophosphate is measured by (a) adding adenosine-5'-phosphosulfate to the feedstock; combining the recovered feedstock containing adenosine-5'-phosphosulfate with an ATP generating polypeptide-ATP converting polypeptide fusion protein such that any inorganic pyrophosphate in the recovered feedstock and the adenosine-5'-phosphosulfate will first react to the form ATP and sulfate and then react with luciferin in the presence of oxygen such that the ATP is consumed to produced AMP, inorganic pyrophosphate, carbon dioxide and light; and (b) measuring the amount of light produced. In a preferred embodiment, the template polymer and ATP generating polypeptide-ATP converting polypeptide fusion protein are immobilized on a solid support.

[0119] The invention also includes a method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) introducing the template nucleic acid polymer into a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added; (b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released; (c) separately recovering each of the feedstocks from the polymerization environment; and (d) measuring the amount of PPi with a thermostable sulfurylase and a luciferase in each of the recovered feedstocks to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer. In one embodiment, the thermostable sulfurylase and the luciferase are joined in a fusion protein. In another embodiment, the thermostable sulfurylase is joined to an affinity tag.

[0120] The invention further provides a method for sequencing a nucleic acid, the method comprising: (a) providing one or more nucleic acid anchor primers; (b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200; (c) annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing an effective amount of a sequencing primer to one or more copies of said covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of a thermostable sulfurylase and a luciferase, thereby determining the sequence of the nucleic acid. In one embodiment, the thermostable sulfurylase and the luciferase are joined in a fusion protein. In another embodiment, the thermostable sulfurylase is joined to an affinity tag.

[0121] Also included in the invention is a method for sequencing a nucleic acid, the method comprising: (a) providing at least one nucleic acid anchor primer; (b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites; (c) annealing a first amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing a second amount of a sequencing primer to one or more copies of the covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, when the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, to yield a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of a thermostable sulfurylase and a luciferase, thereby determining the sequence of the nucleic acid at each reaction site that contains a nucleic acid template. In one embodiment, the thermostable sulfurylase and the luciferase are joined in a fusion protein. In another embodiment, the thermostable sulfurylase is joined to an affinity tag.

[0122] The invention also includes a method of determining the base sequence of a plurality of nucleotides on an array, the method comprising: (a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber,.wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, (b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated

nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (c) detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands through detection of a sequencing byproduct with a thermostable sulfurylase and luciferase, thus indicating that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and (d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and (e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors. In one embodiment, the thermostable sulfurylase and the luciferase are joined in a fusion protein. In another embodiment, the thermostable sulfurylase is joined to an affinity tag.

[0123] The invention will be further illustrated in the following non-limiting examples. There are several abbreviations which will be used in the following examples: FUS stands for fusion gene, S stands for sulfurylase, L stands for luciferase, TL stands for thermostable luciferase, X stands for XhoI, H stands for HindIII, N stands for NotI and B stands for BamHI. For example, FUS-L/S X F means a primer for the fusion gene, luciferase-sulfurylase Xho Forward and so forth. Primers 1 through 6 are for the L or TL to S fusions and primers 7 through 13 are for the S to L or TL fusions.

EXAMPLES

Example 1: Cloning Strategy for Obtaining the Bst Sulfurylase Gene

[0124] Gene specific primers, which incorporated restriction site linkers, were designed based on the sequence for a putative ATP sulfurylase from *Bacillus stearothermophilus* in ERGO, a curated database of genomic DNA made available on the World Wide Web by Integrated Genomics which included the *Bacillus stearothermophilus* Genome Sequencing Project at the University of Oklahoma (NSF Grant #EPS-9550478). The forward primer utilized was 5'-CCC TTC TGC AGC ATG AGC GTA AGC ATC CCG CAT GGC GGC ACA TTG-3' (SEQ ID NO: 7) and the reverse primer used was 5'-CCC GTA AGC TTT TAG CGC GCT GAC GGG GCG ACC GTT TCG CGT TCT TG-3' (SEQ ID NO:8). The reaction mix for PCR amplification contained 5.0 uL 10X polymerase buffer (Clontech, Cat. #8714), 2.0 uL 5 M betaine (Sigma, Cat. #B0300), 1.0 uL dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP), 0.8 uL Advantage 2 polymerase (Clontech, Cat. #8714), 0.2 uL Advantage-HF 2 polymerase (Clontech, Cat. #K1914), 10 pmol forward primer, 10 pmol reverse primer, 100 ng (or less) *Bst* genomic DNA (ATCC, Cat. #12980D), and enough distilled water to make total volume of 50 uL. As little as 1 ng *Bst* genomic DNA was sufficient to yield PCR product. The PCR amplification of Bst ATP sulfurylase gene from genomic DNA consisted of an initial step at 96˚C for 3 min, then 35 cycles of 96˚C for 15 sec, 60˚C for 30 sec, 72˚C for 6 min, a finishing step at 72˚C for 10 min and finally 14˚C until removal. The PCR product was cleaned using QIAquick PCR Purification Kit (QIAGEN).

Example 2: Cloning Strategy for Obtaining the Sulfurylase-Luciferase Fusion Protein

[0125] All chemicals were purchased from Sigma unless noted otherwise. Racemically pure D-luciferin was ordered from Pierce. The assay buffer for measuring ATP sulfurylase and luciferase activities contained *Taq* polymerase. A polymerase chain reaction (PCR)-mediated approach was utilized to link the open reading frames (ORF_S) of luciferase and sulfurylase. The cloning strategy is outlined in Fig. 1. Briefly, it involved the amplification of luciferase and sulfurylase ORFs by PCR, using primers that contain convenient restriction sites (XhoI and HindIII) to clone the fusion gene into an expression vector, in-frame and, the design of a rare restriction site (Not I) at the junction of the two polypeptides so that other versions of luciferase, such as thermostable luciferase (TL), and sulfurylase can be conveniently swapped to obtain either sulfurlyase-luciferase (S-L) or luciferase-sulfurlyase (L-S) fusion proteins. A Not I site was used to fuse the variable heavy chain of antibodies to luciferase to generate a viable fusion protein. These primers were also designed in such a way that the primers that form part of the junction of the two ORFs contain sufficient overlapping regions of nucleotides. For example, the 5' end of FUS-L/S Not R contains deoxynucleotides in an antiparallel orientation that encode the N-terminal 10 amino acids of yeast sulfurylase. Thus, a PCR product generated using this primer would anneal to the 5' end of yeast sulfurylase ORF and would generate the fusion protein, L-S.

[0126] The products in boxes were obtained by PCR as elaborated in Fig.2. As shown in Fig. 3, the PCR products were subjected to electrophoresis. The PCR products were then purified, digested, with Xho I and Hind III and subcloned into Xho I/Hind III digested pRSETA-BCCP. pRSETA-BCCP is a derivative of pRSET A (Invitrogen) in which the sequence between *NheI* and *Bam*HI restriction sites has been replaced by the portion of the biotin carboxyl carrier protein (BCCP) gene from E. *coli* (GenBank accession #M80458) that codes for residues 87-165. The 87- amino acid BCCP domain was obtained by PCR and cloned into the NheI and Barn HI sites of pRSETA to obtain pRSETA-BCCP. The ligated fusion protein and pRSETA-BCCP were transformed into BL21DE3 and TOP10 cells. BL21DE3 cells yielded colonies for L-S and TOP10 cells yielded colonies for TL-S.

[0127] The following list of primers was used to construct the fusion proteins:

PRIMER NO	TITLE	NUCLEIC ACID SEQUENCE	SEQ ID NO
1	FUS-L/S X F	CCCC CTC GAG ATC CAA ATG GAA GAC GCC AAA AAC ATA AAG AAA GGC CC	9
2	FUS-TL/S X F	CCCC CTC GAG ATC CAA ATG GCT GAC AAA AAC ATC CTG TAT GGC CC	10
3	FUS-L/S Not R	TTG TAG AAT ACC ACC GTG AGG AGC AGG CAT AGC GGC CGC CAA TTT GGA CTT TCC GCC CTT CTT GGC C	11
4	FUS-TL/S NotR	TTG TAG AAT ACC ACC GTG AGG AGC AGG CAT AGC GGC CGC ACC GTT GGT GTG TTT CTC GAA CAT C	12
5	FUS-S-Not F	GCG GCC GCT ATG CCT GCT CCT CAC GGT GGT ATT CTA C	13
6	FUS-S-Hind III R	CCCC AAG CTT TTA AAA TAC AAA AAA GCC ATT GTC TTC CAA GAA TAG GAC	14
7	FUS-S/L B F	CCCC GGA TCC ATC CAA ATG CCT GCT CCT CAC GGT GGT ATT CTA CAA GAC	15
8	FUS-S/L R	GGGCCTTTCTTTATGTTTTTGGCGTCTTCCAT AGC GGC CGC AAA TAC AAA AAA GCC ATT GTC	16
9	FUS-L- F	GCG GCC GCT ATG GAA GAC GCC AAA AAC ATA AAG AAA GGC CC	17
10	FUS-L-N-R	CCCC CCA TGG TTA CAA TTT GGA CTT TCC GCC CTT CTT GGC C	18
11	FUS-S/TLR	GG GCC ATA CAG GAT GTT TTT GTC AGC CAT AGC GGC CGC AAA TAC AAA AAA GCC ATT GTC	19
12	FUS-TL-F	GCG GCC GCT ATG GCT GAC AAA AAC ATC CTG TAT GGC CC	20
13	FUS-TL-H-R	CCCC AAG CTT CTA ACC GTT GGT GTG TTT CTC GAA CAT CTG ACG C	21

[0128] These primers were utilized to perform PCR. The following PCR condition was used.

PCR condition

[0129]

96˚C for 3:00; 96˚C for 0:15; 76˚C for 0:30; -1˚C per cycle; 72˚C for 6:00;
For 15 cycles; 96˚C for 0:15; 60˚C for 0:30; 72˚C for 6:00;
For 29 cycles; 72˚C for 10:00;
14˚C forever

Example 3: Cloning of the His6-BCCP Bst ATP Sulfurylase Fusion Protein

[0130] The Bst-affinity tagged fusion construct is a derivative of pRSETA in which the NheI-XhoI fragment has been

replaced by the BCCP domain and the ATP sulfurylase is inserted after the BCCP domain.

[0131] Briefly, the BstSulf PCR product, as described in Example 1, was double-digested with *PstI* and *Hind*III, isolated on a 1% agarose/TAE gel, purified using QIAEXII (QIAGEN) and ligated into the large *Pst*I/*HindIII* fragment of pRSETA-BCCP using the Quick Ligation Kit from NEB according to manufacturer's instructions. As mentioned in Example 2, pRSETA-BCCP is a derivative of pRSET A (Invitrogen) in which the sequence between NheI and *Bam*HI restriction sites has been replaced by the portion of the biotin carboxyl carrier protein (BCCP) gene from *E. coli* (GenBank accession #M80458) that codes for residues 87-165. 2 uL ligation reaction was used to transform 50 uL TOP10 competent cells (Invitrogen) and plated on LB-Ap plates. Sequencing of plasmid insert from ten clones was used to determine the consensus sequence for the ATP sulfurylase gene from ATCC 12980.

[0132] The plasmid pRSETA-BCCP-BstSulf was transformed into the *E. coli* expression host BL21(DE3)pLysS (Novagen) and the induction expression of BstHBSulf was carried out according to the maufacturer's instructions. The cells were harvested and stored as frozen pellets. The pellets were lysed using BugBuster plus Benzonase according to manufacturer's instructions and protein was purified on a 20 mL column packed with Chelating Sepharose Fast Flow (Amersham, Cat. #17-0575-02) and charged with nickel (II). Protein was eluted using a 0-500 mM imidazole gradient. Analysis by SDS-PAGE showed a single band of the correct size.

Example 4: Binding enzymes to beads

[0133] The BCCP domain enables the E. *coli* to add a single biotin molecule onto a specific lysine residue. Hence these fusion proteins can be bound to solid supports that contain streptavidin. TL-S was successfully cloned into a TA vector. 25 μl of MPG-Streptavidin (CPG, Inc.) or Nickel-agarose (Qiagen) were taken in a 1.5 ml tube and placed on a magnet. The supernatant was removed and the beads were resuspended in 25 μg of His6-BCCP-sulfurylase and 75 μg of His6-BCCP-luciferase. To test the fusion protein, 100 μl of dialyzed fusion protein was bound to the 25 μl of beads. The beads were allowed to mix at room temperature for 1hr, washed with assay buffer (25mM Tricine (pH 7.8), 5mM MgAcetate, 1mM DTT, 1mM EDTA, and 1 mg/ml BSA) and assayed for enzyme activities with 1mM PPi, 4mM APS and 300 mM D-luciferin. With the nickel-agarose beads, the EDTA was omitted from the assay buffer.

[0134] As shown in Fig. 4, these fusion proteins displayed activity on both the NTA- Agarose and MPG-SA beads. S:L 1:3 represents sulfurylase and luciferase bound individually to beads in a 1:3 ratio. Ni-Ag and MPG-SA are nickel-agarose and MPG-Streptavidin beads, respectively. PL is Promega luciferase, which does not have a polyhistidine or a biotin tag on it and hence serves as a negative control. Fraction 19 contains the fusion protein and is active on both kinds of beads. This suggests that the fusion protein was synthesized with a poly-histidine tag and a biotin molecule on the BCCP domain of the fusion protein.

EMBODIMENTS OF THE DISCLOSURE

[0135]

1. A fusion protein comprising an ATP generating polypeptide bound to a polypeptide which converts ATP to an entity that is detectable.

2. The fusion protein of embodiment 1 wherein the ATP generating polypeptide is selected. from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

3. The fusion protein of embodiment 2 wherein the ATP sulfurylase is a thermostable sulfurylase comprising the nucleotide sequence of SEQ ID NO:1.

4. The fusion protein of embodiment 3 wherein the nucleotide sequence encodes the polypeptide sequence of SEQ ID NO:2.

5. The fusion protein of embodiment 3 wherein the thermostable sulfurylase is active at room temperature.

6. The fusion protein of embodiment 2 wherein the ATP sulfurylase is from a thermophile.

7. The fusion protein of embodiment 6 wherein the thermophile is a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

8. The fusion protein of embodiment 1 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

9. The fusion protein of embodiment 8 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

10. The fusion protein of embodiment 9 wherein the animal is selected from the group consisting of mammal, rodent, insect, worm, mollusk, reptile, bird and amphibian.

11. The fusion protein of embodiment 9 wherein the plant is selected from the group consisting *of Arabidopsis thaliana, Brassica napus, Allium sativum, Amaranthus caudatus, Hevea brasiliensis, Hordeum vulgare, Lycopersicon esculentum, Nicotiana tabacum, Oryza sativum, Pisum sativum, Populus trichocarpa, Solanum tuberosum, Secale cereale, Sambucus nigra, Ulmus americana* or *Triticum aestfvum.*

12. The fusion protein of embodiment 9 wherein the fungus is *Penicillum chrysogenum, Stachybotrys chartarum, Aspergillus fumigatus, Podospora anserina, Trichoderma reesei* and *Riftia pachyptila.*

13. The fusion protein of embodiment 9 wherein the yeast is *Saccharomyces cerevisiae, Candida tropicalis, Candida lypolitica, Candida utilis, Kluyveromyces lactis, Schizosaccharomyces pombe, Yarrowia lipolytica, Candida spp., Pichia spp.* and *Hansenula spp..*

14. The fusion protein of embodiment 8 wherein the prokaryote is bacteria or archaea.

15. The fusion protein of embodiment 14 wherein the bacteria selected from the group consisting of *E. coli, B. subtilis, Streptococcus gordonii,* flavobacteria and green sulfur bacteria.

16. The fusion protein of embodiment 14 wherein the archaea is selected from the group consisting of *Sulfolobus, Thermococcus, Methanobacterium, Halococcus, Halobacterium* and *Methanococcus jannaschii.*

17. The fusion protein of embodiment 1 wherein the detectable entity is selected from the group consisting of chemiluminescence, bioluminescence and fluorescence.

18. The fusion protein of embodiment1 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

19. The fusion protein of embodiment 18 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

20. The fusion protein of embodiment 1 which further comprises an affinity tag.

21. The fusion protein of embodiment 20 wherein the affinity tag is selected from the group consisting of N-terminal poly-histidine, BCCP, protein A, glutathione S transferase, substance P and streptavidin binding peptide.

22. The fusion protein of embodiment 1 wherein the polypeptides are joined by a linker.

23. The fusion protein of embodiment 22 wherein the linker is an ala-ala-ala linker.

24. The fusion protein of embodiment 1 wherein the ATP generating polypeptide is N-terminal to the ATP converting polypeptide.

25. The fusion protein of embodiment 1 wherein the ATP converting polypeptide is N-terminal to the ATP generating polypeptide.

26. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3 and 5.

27. An isolated polypeptide comprising an amino acid sequence selected from the group consising of SEQ ID NOs: 2, 4 and 6.

28. A fusion protein comprising a sulfurylase polypeptide bound to a luciferase polypeptide and at least one affinity tag.

29. The fusion protein of embodiment 28 wherein the fusion protein comprises the sequence of SEQ ID NO:4.

30. The fusion protein of embodiment 28 wherein the fusion protein is encoded by a nucleic acid comprising the sequence of SEQ ID. N0:3.

31. A fusion protein comprising a thermostable sulfurylase bound to at least one affinity tag.

32. The fusion protein of embodiment 31 wherein the fusion protein comprises the sequence of SEQ ID NO: 6.

33. The fusion protein of embodiment 31 wherein the fusion protein is encoded by a nucleic acid comprising the sequence of SEQ ID NO:5.

34. A recombinant polynucleotide that comprises a coding sequence for a fusion protein having an ATP generating polypeptide sequence and an ATP converting polypeptide sequence.

35. The recombinant polynucleotide sequence of embodiment34 wherein the ATP generating polypeptide is ATP sulfurylase.

36. The recombinant polynucleotide sequence of embodiment 34 wherein the ATP converting polypeptide is luciferase.

37. The recombinant polynucleotide of embodiment 34 wherein the ATP generating polypeptide is N-terminal to the ATP converting polypeptide.

38. The recombinant polynucleotide of embodiment 34 wherein the ATP converting polypeptide is N-terminal to the ATP generating polypeptide.

39. An expression vector for expressing a fusion protein, said vector comprising a coding sequence for a fusion protein having: (i) a regulatory sequence, (ii) a first polypeptide sequence of an ATP generating polypeptide and (iii) a second polypeptide sequence that converts ATP to an entity which is detectable.

40. The expression vector of claim 39 wherein the vector further comprises an affinity tag.

41. The expression vector of embodiment 39 wherein the ATP generating polypeptide is ATP sulfurylase.

42. The expression vector of embodiment 39 wherein the ATP converting polypeptide is luciferase.

43. The expression vector of embodiment 39 wherein the regulatory element is an enhancer or a promoter.

44. The expression vector of embodiment 43 wherein the promoter is a constitutive promoter or an inducible promoter.

45. A transformed host cell which contains the expression vector of embodiment 39.

46. The transformed host cell of embodiment 45 wherein the host cell is a eukaryotic cell.

47. The transformed host cell of embodiment 46 wherein the eukaryotic cell is human, rat or mouse.

48. The transformed host cell of embodiment 45 wherein the host cell is a prokaryotic cell.

49. The transformed host cell of embodiment 48 wherein the prokaryotic cell is bacteria.

50. A purified fusion protein expressed by cells transformed with an expression vector of embodiment 39.

51. The fusion protein of embodiment 1 bound to a mobile support.

52. The fusion protein of embodiment 51 wherein the fusion protein is attached by a covalent or non-covalent

interaction.

53. The fusion protein of embodiment 52 wherein the fusion protein is attached by a linkage selected from the group consisting of a metal, a CO^{2}+-hexahistidine complex, a Ni^{2+}-hexahistidine complex, a biotin binding protein, a glutathione S-transferase/glutathione complex, a monoclonal antibody/antigen complex, a maltose binding protein/maltose complex and pluronic coupling.

54. The fusion protein of embodiment 53 wherein the biotin binding protein is selected from the group consisting of NEUTRAVIDINT™ modified avidin, streptavidin and avidin.

55. The fusion protein of embodiment 51 wherein the mobile support is selected from the group consisting of a bead, optical fiber and glass surface.

56. The fusion protein of embodiment 55 wherein the bead is a nickel-agarose bead or a MPG-Streptavidin bead.

57. The fusion protein of embodiment51 wherein the fusion protein is bound to the mobile support in a 1:3 ratio of protein to mobile support.

58. The fusion protein of embodiment51 wherein the fusion protein is a sulfurylase-luciferase fusion protein.

59. A method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising:

(a) introducing the template nucleic acid polymer into a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added;
(b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released;
(c) separately recovering each of the feedstocks from the polymerization environment; and
(d) measuring the amount of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein in each of the recovered feedstocks to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer.

60. The method of embodiment 59 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

61. The method of embodiment 60 wherein the ATP sulfurylase is a thermostable sulfurylase.

62. The method of embodiment 60 wherein the ATP sulfurylase is from a thermophile.

63. The method of embodiment 62 wherein the thermophile is a thermophilic bacteria selected from the group consisting *of Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

64. The method of embodiment 59 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

65. The method of embodiment 64 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

66. The method of embodiment 59 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

67. The method of embodiment 66 wherein the luciferase is selected from the group consisting of *Photinus pyralis,*

Pyroplorus plaglophihalamus (Coleoptera), Luciola cruciata and *Luciola lateralis.*

68. The method of embodiment 59 which further comprises an affinity tag.

69. A method according to embodiment 59, wherein the amount of inorganic pyrophosphate is measured by

(a) adding adenosine-5'-phosphosulfate to the feedstock;
(b) combining the recovered feedstock containing adenosine-5'-phosphosulfate with an ATP generating polypeptide-ATP converting polypeptide fusion protein such that any inorganic pyrophosphate in the recovered feedstock and the adenosine-5'-phosphosulfate will react to the form ATP and sulfate;
(c) combining the ATP, sulfate, and said fusion protein-containing feedstock with luciferin in the presence of oxygen such that the ATP is consumed to produced AMP, inorganic pyrophosphate, carbon dioxide and light; and
(d) measuring the amount of light produced.

70. The method of embodiment 69 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

71. The method of embodiment 70 wherein the ATP sulfurylase is a thermostable sulfurylase.

72. The method of embodiment 70 wherein the ATP sulfurylase is from a thermophile.

73. The method of embodiment 72 wherein the thermophile is a thermophilic bacteria selected from the group consisting *of Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

74. The method of embodiment 69 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

75. The method of embodiment 74 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

76. The method embodiment 69 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

77. The method of embodiment 76 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

78. The method of embodiment 69 which further comprises an affinity tag.

79. The method according to embodiment 59 wherein each feedstock comprises adenosine-5'-phosphosulfate and luciferin in addition to the selected nucleotide base, and the amount of inorganic pyrophosphate is determined by reacting the inorganic pyrophosphate is determined by reacting the inorganic pyrophosphate-containing feedstock with an ATP generating polypeptide-ATP converting polypeptide fusion protein thereby producing light in an amount proportional to the amount of inorganic pyrophosphate, and measuring the amount of light produced.

80. The method of embodiment 79 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

81. The method of embodiment 80 wherein the ATP sulfurylase is a thermostable sulfurylase.

82. The method of embodiment 80 wherein the ATP sulfurylase is from a thermophile.

83. The method of embodiment 82 wherein the thermophile is a thermophilic bacteria selected from the group consisting *of Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysoge-*

num, Sulfolobus solfataricus and *Thermomonospora fusca.*

84. The method of embodiment 79 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

85. The method of embodiment 84 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

86. The method of embodiment 79 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

87. The method of embodiment 86 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

88. The method of embodiment 79 which further comprises an affinity tag.

89. A method for sequencing a nucleic acid, the method comprising:

(a) providing one or more nucleic acid anchor primers;
(b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm;
(c) annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex;
(d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template;
(e) annealing an effective amount of a sequencing primer to one or more copies of said covalently linked complementary nucleic acid;
(f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and
(g) identifying the sequencing reaction byproduct with the use of an ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the sequence of the nucleic acid.

90. The method of embodiment 89 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

91. The method of embodiment 90 wherein the ATP sulfurylase is a thermostable sulfurylase.

92. The method of embodiment 90 wherein the ATP sulfurylase is from a thermophile.

93. The method of embodiment 92 wherein the thermophile is a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

94. The method of embodiment 89 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

95. The method of embodiment 94 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

96. The method of embodiment 89 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

97. The method of embodiment 96 wherein the luciferase is selected from the group consisting of *Photinus pyralis,*

Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata and *Luciola lateralis.*

98. The method of embodiment 89 which further comprises an affinity tag.

99. A method for sequencing a nucleic acid, the method comprising:

(a) providing at least one nucleic acid anchor primer;
(b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites;
(c) annealing a first amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex;
(d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template;
(e) annealing a second amount of a sequencing primer to one or more copies of the covalently linked complementary nucleic acid;
(f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, when the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, to yield a sequencing reaction byproduct; and
(g) identifying the sequencing reaction byproduct with the use of an ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the sequence of the nucleic acid at each reaction site that contains a nucleic acid template.

100. The method of embodiment 99 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

101. The method of embodiment 100 wherein the ATP sulfurylase is a thermostable sulfurylase.

102. The method of embodiment 100 wherein the ATP sulfurylase is from a thermophile.

103. The method of embodiment 102 wherein the thermophile is a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

104. The method of embodiment 99 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

105. The method of embodiment 104 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

106. The method of embodiment 99 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

107. The method of embodiment 106 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

108. The method of embodiment 99 which further comprises an affinity tag.

109. A method of determining the base sequence of a plurality of nucleotides on an array, the method comprising:

(a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm,
(b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least

one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates;
(c) detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands through detection of a sequencing byproduct with an ATP generating polypeptide-ATP converting polypeptide fusion protein, thus indicating that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and
(d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and
(e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

110. The method of embodiment 109 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

111. The method of embodiment 110 wherein the ATP sulfurylase is a thermostable sulfurylase.

112. The method of embodiment 110 wherein the ATP sulfurylase is from a thermophile.

113. The method of claim 112 wherein the thermophile is a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

114. The method of embodiment 109 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

115. The method of embodiment 114 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

116. The method of embodiment 109 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

117. The method of embodiment 116 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

118. The method of embodiment 109 which further comprises an affinity tag.

119. A method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising:

(a) introducing a plurality of template nucleic acid polymers into a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, each reaction chamber having a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added;
(b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released;
(c) detecting the formation of inorganic pyrophosphate with an ATP generating polypeptide-ATP converting polypeptide fusion protein to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer.

120. The method of embodiment 119 wherein the ATP generating polypeptide is selected from the group consisting

of ATP sulfurylase, hydrolase and ATP synthase.

121. The method of embodiment 120 wherein the ATP sulfurylase is a thermostable sulfiurylase.

122. The method of embodiment 120 wherein the ATP sulfurylase is from a thermophile.

123. The method of embodiment 122 wherein the thermophile is a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

124. The method of embodiment 119 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

125. The method of embodiment 124 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

126. The method of embodiment 119 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

127. The method of embodiment 126 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

128. The method of embodiment 119 which further comprises an affinity tag.

129. A method of identifying the base in a target position in a DNA sequence of sample DNA, wherein:

(a) sample DNA is disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, said DNA being rendered single stranded either before or after being disposed in the reaction chambers,
(b) an extension primer is provided which hybridizes to said immobilized single-stranded DNA at a position immediately adjacent to said target position;
(c) said immobilized single-stranded DNA is subjected to a polymerase reaction in the presence of a predetermined nucleotide triphosphate, wherein if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer then a sequencing reaction byproduct is formed; and
(d) identifying the sequencing reaction byproduct with an ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the nucleotide complementary to the base at said target position.

130. The method of embodiment 129 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

131. The method of embodiment 130 wherein the ATP sulfurylase is a thermostable sulfurylase.

132. The method of embodiment 130 wherein the ATP sulfiuylase is from a thermophile.

133. The method of embodiment 132 wherein the thermophile is a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

134. The method of embodiment 129 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

135. The method of embodiment 134 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

136. The method of embodiment 129 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

137. The method of embodiment 136 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

138. The method of embodiment 129 which further comprises an affinity tag.

139. A method of identifying a base at a target position in a sample DNA sequence comprising:

(a) providing sample DNA disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, said DNA being rendered single stranded either before or after being disposed in the reaction chambers;
(b) providing an extension primer which hybridizes to the sample DNA immediately adjacent to the target position;
(c) subjecting the sample DNA sequence and the extension primer to a polymerase reaction in the presence of a nucleotide triphosphate whereby the nucleotide triphosphate will only become incorporated and release pyrophosphate (PPi) if it is complementary to the base in the target position, said nucleotide triphosphate being added either to separate aliquots of sample-primer mixture or successively to the same sample-primer mixture; and
(d) detecting the release of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein to indicate which nucleotide is incorporated.

140. The method of embodiment 139 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

141. The method of embodiment 140 wherein the ATP sulfurylase is a thermostable sulfurylase.

142. The method of embodiment 140 wherein the ATP sulfruylase is from a thermophile.

143. The method of embodiment 142 wherein the thermophile is a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorarcs, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

144. The method of embodiment 139 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

145. The method embodiment 144 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

146. The method of embodiment 139 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

147. The method embodiment 146 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

148. The method of embodiment 139 which further comprises an affinity tag.

149. A method of identifying a base at a target position in a single-stranded sample DNA sequence, the method comprising:

(a) providing an extension primer which hybridizes to sample DNA immediately adjacent to the target position, said sample DNA disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 um, said DNA being rendered single stranded either before or after being disposed in the reaction chambers;
(b) subjecting the sample DNA and extension primer to a polymerase reaction in the presence of a predetermined deoxynucleotide or dideoxynucleotide whereby the deoxynucleotide or dideoxynucleotide will only become

incorporated and release pyrophosphate (PPi) if it is complementary to the base in the target position, said predetermined deoxynucleotides or dideoxynucleotides being added either to separate aliquots of sample-primer mixture or successively to the same sample-primer mixture,
(c) detecting any release of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein to indicate which deoxynucleotide or dideoxynucleotide is incorporated;

characterized in that, the PPi-detection enzyme(s) are included in the polymerase reaction step and in that in place of deoxy- or dideoxy adenosine triphosphate (ATP) a dATP or ddATP analogue is used which is capable of acting as a substrate for a polymerase but incapable of acting as a substrate for a said PPi-detection enzyme.

150. The method of embodiment 149 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

151. The method embodiment 150 wherein the ATP sulfurylase is a thermostable sulfurylase.

152. The method of embodiment 150 wherein the ATP sulfurylase is from a thermophile.

153. The method of embodiment 152 wherein the thermophile is a thermophilic bacteria selected from the group consisting *of Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

154. The fusion protein of embodiment 149 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

155. The method of embodiment 154 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

156. The method of embodiment 149 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

157. The method of embodiment 156 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

158. The method of embodiment 149 which further comprises an affinity tag.

159. A method of determining the base sequence of a plurality of nucleotides on an array, the method comprising:

(a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm,
(b) converting PPi into light with a an ATP generating polypeptide-ATP converting polypeptide fusion protein;
(c) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device;
(d) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions;
(e) determining a light intensity for each of said discrete regions from the corresponding electrical signal;
(f) recording the variations of said electrical signals with time.

160. The method of embodiment 159 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

161. The method of embodiment 160 wherein the ATP sulfurylase is a thermostable sulfurylase.

162. The method of embodiment 160 wherein the ATP sulfurylase is from a thermophile.

163. The method of embodiment 162 wherein the thermophile is a thermophilic bacteria selected from the group

consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

164. The method of embodiment 159 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

165. The method of embodiment 164 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

166. The method of embodiment 159 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

167. The method of embodiment wherein the luciferase is selected from group consisting 167. The method of embodiment 166 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

168. The method of embodiment 159 which further comprises an affinity tag.

169. Method for sequencing a nucleic acid, the method comprising:

(a) providing one or more nucleic acid anchor primers;
(b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm;
(c) converting PPi into a detectable entity with the use of an ATP generating polypeptide-ATP converting polypeptide fusion protein;
(d) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device;
(e) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions;
(f) determining a light intensity for each of said discrete regions from the corresponding electrical signal;
(g) recording the variations of said electrical signals with time.

170. The method of embodiment 169 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

171. The method of embodiment 170 wherein the ATP sulfurylase is a thermostable sulfurylase.

172. The method of embodiment 170 wherein the ATP sulfurylase is from thermophile.

173. The method of embodiment 172 wherein the thermophile is a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus filgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

174. The method of embodiment 169 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

175. The method of embodiment 174 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

176. The method of embodiment 169 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

177. The method of embodiment 176 wherein the luciferase is selected from the group consisting of *Photinus pyralis,*

Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata and *Luciola lateralis.*

178. The method of embodiment 169 which further comprises an affinity tag.

179. A method for sequencing a nucleic acid, the method comprising:

(a) providing at least one nucleic acid anchor primer;
(b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites;
(c) converting PPi into a detectable entity with an ATP generating polypeptide-ATP converting polypeptide fusion protein;
(d) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device;
(e) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions;
(f) determining a light intensity for each of said discrete regions from the corresponding electrical signal;
(g) recording the variations of said electrical signals with time.

180. The method of embodiment 179 wherein the ATP generating polypeptide is selected from the group consisting of ATP sulfurylase, hydrolase and ATP synthase.

181. The method of embodiment 180 wherein the ATP sulfurylase is a thermostable sulfurylase.

182. The method of embodiment 180 wherein the ATP sulfurylase is from a thermophile.

183. The method of embodiment 182 wherein the thermophile is a thermophilic bacteria selected from the group consisting *of Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcusfiiriosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobusfulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenurn, Sulfolobus solfataricus* and *Thermomonospora fusca.*

184. The method of embodiment 179 wherein the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote.

185. The method of embodiment 184 wherein the eukaryote is selected from the group consisting of animal, plant, fungus and yeast.

186. The method of embodiment 179 wherein the ATP converting polypeptide is selected from the group consisting of luciferase, ecto-nucleoside diphosphate kinase and ATPase.

187. The methods of embodiment 186 wherein the luciferase is selected from the group consisting of *Photinus pyralis, Pyroplorus plagiophihalamus (Coleoptera), Luciola cruciata* and *Luciola lateralis.*

188. The method of embodiment 179 which further comprises an affinity tag.

189. A kit comprising a sulfurylase-luciferase fusion protein expression vector as claimed in embodiment 39.

190. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) a mature form of an amino acid sequence of SEQ ID NO: 2;
(b) a variant of a mature form of an amino acid sequence of SEQ ID NO: 2;
(c) an amino acid sequence of SEQ ID NO: 2;
(d) a variant of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 4% of amino acid residues from said amino acid sequence; and
(e) an amino acid sequence of (a), (b), (c) or (d) further containing one or more conservative amino acid substitutions.

191. The polypeptide of embodiment 190 wherein said polypeptide comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence of SEQ ID NO: 2.

192. The polypeptide of embodiment 190 wherein the amino acid sequence of said variant comprises one or more conservative amino acid substitution.

193. An isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) a mature form of an amino acid sequence of SEQ ID NO: 2;
(b) a variant of a mature form of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 4% of the amino acid residues from the amino acid sequence of said mature form;
(c) an amino acid sequence of SEQ ID NO: 2;
(d) a variant of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 4% of amino acid residues from said amino acid sequence;
(e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising an amino acid sequence of SEQ ID NO: 2, or a variant of said polypeptide, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 4% of amino acid residues from said amino acid sequence; and
(f) a nucleic acid molecule comprising the complement of (a), (b), (c), (d) or (e).

194. The nucleic acid molecule of embodiment 193 wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally-occurring allelic nucleic acid variant.

195. The nucleic acid molecule of embodiment 193 wherein the nucleic acid molecule encodes a polypeptide comprising the amino acid sequence of a naturally-occurring polypeptide variant.

196. The nucleic acid molecule of embodiment 193 wherein the nucleic acid molecule comprises nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 11% of the nucleotides in the coding sequence in said first nucleotide sequence differ from said coding sequence;
(b) an isolated second polynucleotide that is a complement of the first polynucleotide; and
(c) a nucleic acid fragment of (a) or (b).

197. A vector comprising the nucleic acid molecule of embodiment 196.

198. The vector of embodiment 197, further comprising a promoter operably-linked to said nucleic acid molecule.

199. A cell comprising the vector of embodiment 197.

200. An antibody that binds immunospecifically to the polypeptide of embodiment 190.

201. A method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising:

(a) introducing the template nucleic acid polymer into a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added;
(b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released;
(c) separately recovering each of the feedstocks from the polymerization environment; and
(d) measuring the amount of PPi with a thermostable sulfurylase and a luciferase in each of the recovered

feedstocks to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer.

202. The method of embodiment 201 wherein said thermostable sulfurylase comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence of SEQ ID NO: 2.

203. The method of embodiment 201 wherein the thermostable sulfurylase is derived from a thermophilic bacteria selected from the group consisting of *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

204. The method of embodiment 201 wherein the thermostable sulfurylase and the luciferase are joined in a fusion protein.

205. The method of embodiment 201 wherein the thermostable sulfurylase is joined to an affinity tag.

206. A method for sequencing a nucleic acid, the method comprising:

(a) providing one or more nucleic acid anchor primers;
(b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200;
(c) annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex;
(d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template;
(e) annealing an effective amount of a sequencing primer to one or more copies of said covalently linked complementary nucleic acid;
(f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and
(g) identifying the sequencing reaction byproduct with the use of a thermostable sulfurylase and a luciferase, thereby determining the sequence of the nucleic acid.

207. The method of embodiment 206 wherein said thermostable sulfurylase comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence of SEQ ID NO: 2.

208. The method of embodiment 206 wherein the thermostable sulfurylase is derived from a thermophilic bacteria selected from the group consisting of Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, *Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

209. The method of embodiment 206 wherein the thermostable sulfurylase and the luciferase are joined in a fusion protein.

210. The method of embodiment 206 wherein the thermostable sulfurylase is joined to an affinity tag.

211. A method for sequencing a nucleic acid, the method comprising:

(a) providing at least one nucleic acid anchor primer;
(b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites;
(c) annealing a first amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex;
(d) combining the primed anchor primer-circular template complex with a polymerase to form an extended

anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template;
(e) annealing a second amount of a sequencing primer to one or more copies of the covalently linked complementary nucleic acid;
(f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, when the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, to yield a sequencing reaction byproduct; and
(g) identifying the sequencing reaction byproduct with the use of a thermostable sulfurylase and a luciferase, thereby determining the sequence of the nucleic acid at each reaction site that contains a nucleic acid template.

212. The method of embodiment 211 wherein said thermostable sulfurylase comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence of SEQ ID NO: 2.

213. The method of embodiment 211 wherein the thermostable sulfurylase is derived from a thermophilic bacteria selected from the group consisting of Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, *Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

214. The method of embodiment 211 wherein the thermostable sulfurylase and the luciferase are joined in a fusion protein.

215. The method of embodiment 211 wherein the thermostable sulfurylase is joined to an affinity tag.

216. A method of determining the base sequence of a plurality of nucleotides on an array, the method comprising:

(a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm,
(b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates;
(c) detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands through detection of a sequencing byproduct with a thermostable sulfurylase and luciferase, thus indicating that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-tripliosphate precursor; and
(d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and
(e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

217. The method of embodiment 216 wherein said thermostable sulfurylase comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence of SEQ ID NO: 2.

218. The method of embodiment 216 wherein the thermostable sulfurylase is derived from a thermophilic bacteria selected from the group consisting of Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, *Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

219. The method of embodiment 216 wherein the thermostable sulfurylase and the luciferase are joined in a fusion protein.

220. The method of embodiment 216 wherein the thermostable sulfurylase is joined to an affinity tag.

SEQUENCE LISTING

<110> Srinivasan, Maithreyan
Reifler, Michael

<120> Novel Sulfurylase-Luciferase Fusion Proteins and Thermostable Sulfurylase

<130> 21465-504CIP1

<140> 10/154,515
<141> 2002-05-23

<150> 60/335,949
<151> 2001-10-30

<150> 10/122,706
<151> 2002-04-11

<160> 31

<170> PatentIn Ver. 2.1

<210> 1
<211> 1247
<212> DNA
<213> Bacillus stearothermophilus

<400> 1
gttatgaaca tgagtttgag cattccgcat ggcggcacat tgatcaaccg ttggaatccg 60
gattacccaa tcgatgaagc aacgaaaacg atcgagctgt ccaaagccga actaagcgac 120
cttgagctga tcggcacagg cgcctacagc ccgctcaccg ggtttttaac gaaagccgat 180
tacgatgcgg tcgtagaaac gatgcgcctc gctgatggca ctgtctggag cattccgatc 240
acgctggcgg tgacggaaga aaaagcgagt gaactcactg tcggcgacaa agcgaaactc 300
gtttatggcg gcgacgtcta cggcgtcatt gaaatcgccg atatttaccg cccggataaa 360
acgaaagaag ccaagctcgt ctataaaacc gatgaactcg ctcacccggg cgtgcgcaag 420
ctgtttgaaa aaccagatgt gtacgtcggc ggagcggtta cgctcgtcaa acggaccgac 480
aaaggccagt ttgctccgtt ttatttcgat ccggccgaaa cgcggaaacg atttgccgaa 540
ctcggctgga ataccgtcgt cggcttccaa acacgcaacc cggttcaccg cgcccatgaa 600
tacattcaaa aatgcgcgct tgaaatcgtg gacggcttgt ttttaaaccc gctcgtcggc 660
gaaacgaaag cggacgatat tccggccgac atccggatgg aaagctatca agtgctgctg 720
gaaaactatt atccgaaaga ccgcgttttc ttgggcgtct tccaagctgc gatgcgctat 780
gccggtccgc gcgaagcgat tttccatgcc atggtgcgga aaaacttcgg ctgcacgcac 840
ttcatcgtcg gccgcgacca tgcgggcgtc ggcaactatt acggcacgta tgatgcgcaa 900
aaaatcttct cgaactttac agccgaagag cttggcatta caccgctctt tttcgaacac 960
agcttttatt gcacgaaatg cgaaggcatg gcatcgacga aaacatgccc gcacgacgca 1020
caatatcacg ttgtcctttc tggcacgaaa gtccgtgaaa tgttgcgtaa cggccaagtg 1080
ccgccgagca cattcagccg tccggaagtg gccgccgttt tgatcaaagg gctgcaagaa 1140
cgcgaaacgg tcaccccgtc gacacgctaa aggaggagcg agatgagcac gaatatcgtt 1200
tggcatcata catcggtgac aaaagaagat cgccgccaac gcaacgg 1247

<210> 2
<211> 386
<212> PRT
<213> Bacillus stearothermophilus

<400> 2
Met Ser Leu Ser Ile Pro His Gly Gly Thr Leu Ile Asn Arg Trp Asn
1 5 10 15

Pro Asp Tyr Pro Ile Asp Glu Ala Thr Lys Thr Ile Glu Leu Ser Lys
20 25 30

Ala Glu Leu Ser Asp Leu Glu Leu Ile Gly Thr Gly Ala Tyr Ser Pro
35 40 45

Leu Thr Gly Phe Leu Thr Lys Ala Asp Tyr Asp Ala Val Val Glu Thr
50 55 60

Met Arg Leu Ala Asp Gly Thr Val Trp Ser Ile Pro Ile Thr Leu Ala

65 70 75 80

Val Thr Glu Glu Lys Ala Ser Glu Leu Thr Val Gly Asp Lys Ala Lys
85 90 95

Leu Val Tyr Gly Gly Asp Val Tyr Gly Val Ile Glu Ile Ala Asp Ile
100 105 110

Tyr Arg Pro Asp Lys Thr Lys Glu Ala Lys Leu Val Tyr Lys Thr Asp
115 120 125

Glu Leu Ala His Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val
130 135 140

Tyr Val Gly Gly Ala Val Thr Leu Val Lys Arg Thr Asp Lys Gly Gln
145 150 155 160

Phe Ala Pro Phe Tyr Phe Asp Pro Ala Glu Thr Arg Lys Arg Phe Ala
165 170 175

Glu Leu Gly Trp Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val
180 185 190

His Arg Ala His Glu Tyr Ile Gln Lys Cys Ala Leu Glu Ile Val Asp
195 200 205

Gly Leu Phe Leu Asn Pro Leu Val Gly Glu Thr Lys Ala Asp Asp Ile
210 215 220

Pro Ala Asp Ile Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr
225 230 235 240

Tyr Pro Lys Asp Arg Val Phe Leu Gly Val Phe Gln Ala Ala Met Arg
245 250 255

Tyr Ala Gly Pro Arg Glu Ala Ile Phe His Ala Met Val Arg Lys Asn
260 265 270

Phe Gly Cys Thr His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly
275 280 285

Asn Tyr Tyr Gly Thr Tyr Asp Ala Gln Lys Ile Phe Ser Asn Phe Thr
290 295 300

Ala Glu Glu Leu Gly Ile Thr Pro Leu Phe Phe Glu His Ser Phe Tyr
305 310 315 320

Cys Thr Lys Cys Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp
325 330 335

Ala Gln Tyr His Val Val Leu Ser Gly Thr Lys Val Arg Glu Met Leu
340 345 350

Arg Asn Gly Gln Val Pro Pro Ser Thr Phe Ser Arg Pro Glu Val Ala
355 360 365

Ala Val Leu Ile Lys Gly Leu Gln Glu Arg Glu Thr Val Thr Pro Ser
370 375 380

Thr Arg
385

<210> 3
<211> 3519
<212> DNA
<213> Escherichia coli

<400> 3
atgcggggtt ctcatcatca tcatcatcat ggtatggcta gcatggaagc gccagcagca 60
gcggaaatca gtggtcacat cgtacgttcc ccgatggttg gtactttcta ccgcacccca 120

```
agcccggacg caaaagcgtt catcgaagtg ggtcagaaag tcaacgtggg cgataccctg 180
tgcatcgttg aagccatgaa aatgatgaac cagatcgaag cggacaaatc cggtaccgtg 240
aaagcaattc tggtcgaaag tggacaaccg gtagaatttg acgagccgct ggtcgtcatc 300
gagggatccg agctcgagat ccaaatggaa gacgccaaaa acataaagaa aggcccggcg 360
ccattctatc ctctagagga tggaaccgct ggagagcaac tgcataaggc tatgaagaga 420
tacgccctgg ttcctggaac aattgctttt acagatgcac atatcgaggt gaacatcacg 480
tacgcggaat acttcgaaat gtccgttcgg ttggcagaag ctatgaaacg atatgggctg 540
aatacaaatc acagaatcgt cgtatgcagt gaaaactctc ttcaattctt tatgccggtg 600
ttgggcgcgt tatttatcgg agttgcagtt gcgcccgcga acgacattta taatgaacgt 660
gaattgctca acagtatgaa catttcgcag cctaccgtag tgtttgtttc caaaaagggg 720
ttgcaaaaaa ttttgaacgt gcaaaaaaaa ttaccaataa tccagaaaat tattatcatg 780
gattctaaaa cggattacca gggatttcag tcgatgtaca cgttcgtcac atctcatcta 840
cctcccggtt ttaatgaata cgattttgta ccagagtcct ttgatcgtga caaaacaatt 900
gcactgataa tgaattcctc tggatctact gggttaccta agggtgtggc ccttccgcat 960
agaactgcct gcgtcagatt ctcgcatgcc agagatccta tttttggcaa tcaaatcatt 1020
ccggatactg cgattttaag tgttgttcca ttccatcacg gttttggaat gtttactaca 1080
ctcggatatt tgatatgtgg atttcgagtc gtcttaatgt atagatttga agaagagctg 1140
tttttacgat cccttcagga ttacaaaatt caaagtgcgt tgctagtacc aaccctattt 1200
tcattcttcg ccaaaagcac tctgattgac aaatacgatt tatctaattt acacgaaatt 1260
gcttctgggg gcgcacctct ttcgaaagaa gtcggggaag cggttgcaaa acgcttccat 1320
cttccaggga tacgacaagg atatgggctc actgagacta catcagctat tctgattaca 1380
cccgaggggg atgataaacc gggcgcggtc ggtaaagttg ttccattttt tgaagcgaag 1440
gttgtggatc tggataccgg gaaaacgctg ggcgttaatc agagaggcga attatgtgtc 1500
agaggaccta tgattatgtc cggttatgta aacaatccgg aagcgaccaa cgccttgatt 1560
gacaaggatg gatggctaca ttctggagac atagcttact gggacgaaga cgaacacttc 1620
ttcatagttg accgcttgaa gtctttaatt aaatacaaag gatatcaggt ggcccccgct 1680
gaattggaat cgatattgtt acaacacccc aacatcttcg acgcgggcgt ggcaggtctt 1740
cccgacgatg acgccggtga acttcccgcc gccgttgttg ttttggagca cggaaagacg 1800
atgacggaaa aagagatcgt ggattacgtc gccagtcaag taacaaccgc gaaaaagttg 1860
cgcggaggag ttgtgtttgt ggacgaagta ccgaaaggtc ttaccggaaa actcgacgca 1920
agaaaaatca gagagatcct cataaaggcc aagaagggcg gaaagtccaa attggcggcc 1980
gctatgcctg ctcctcacgg tggtattcta caagacttga ttgctagaga tgcgttaaag 2040
aagaatgaat tgttatctga agcgcaatct tcggacattt tagtatggaa cttgactcct 2100
agacaactat gtgatattga attgattcta aatggtgggt tttctcctct gactgggttt 2160
ttgaacgaaa acgattactc ctctgttgtt acagattcga gattagcaga cggcacattg 2220
tggaccatcc ctattacatt agatgttgat gaagcatttg ctaaccaaat taaaccagac 2280
acaagaattg ccctttttcca agatgatgaa attcctattg ctatacttac tgtccaggat 2340
gtttacaagc caaacaaaac tatcgaagcc gaaaaagtct tcagaggtga cccagaacat 2400
ccagccatta gctatttatt taacgttgcc ggtgattatt acgtcggcgg ttctttagaa 2460
gcgattcaat tacctcaaca ttatgactat ccaggtttgc gtaagacacc tgcccaacta 2520
agacttgaat tccaatcaag acaatgggac cgtgtcgtag ctttccaaac tcgtaatcca 2580
atgcatagag cccacaggga gttgactgtg agagccgcca gagaagctaa tgctaaggtg 2640
ctgatccatc cagttgttgg actaaccaaa ccaggtgata tagaccatca cactcgtgtt 2700
cgtgtctacc aggaaattat taagcgttat cctaatggta ttgctttctt atccctgttg 2760
ccattagcaa tgagaatgag tggtgataga gaagccgtat ggcatgctat tattagaaag 2820
aattatggtg cctcccactt cattgttggt agagaccatg cgggcccagg taagaactcc 2880
aagggtgttg atttctacgg tccatacgat gctcaagaat tggtcgaatc ctacaagcat 2940
gaactggaca ttgaagttgt tccattcaga atggtcactt atttgccaga cgaagaccgt 3000
tatgctccaa ttgatcaaat tgacaccaca aagacgagaa ccttgaacat ttcaggtaca 3060
gagttgagac gccgtttaag agttggtggt gagattcctg aatggttctc atatcctgaa 3120
gtggttaaaa tcctaagaga atccaaccca ccaagaccaa aacaaggttt ttcaattgtt 3180
ttaggtaatt cattaaccgt ttctcgtgag caattatcca ttgctttgtt gtcaacattc 3240
ttgcaattcg gtggtggcag gtattacaag atctttgaac acaataataa gacagagtta 3300
ctatctttga ttcaagattt cattggttct ggtagtggac taattattcc aaatcaatgg 3360
gaagatgaca aggactctgt tgttggcaag caaaacgttt acttattaga tacctcaagc 3420
tcagccgata ttcagctaga gtcagcggat gaacctattt cacatattgt acaaaaagtt 3480
gtcctattct tggaagacaa tggcttttt gtattttaa                        3519
```

<210> 4
<211> 1172
<212> PRT
<213> Escherichia coli

<400> 4

```
Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Glu
  1               5                  10                  15

Ala Pro Ala Ala Ala Glu Ile Ser Gly His Ile Val Arg Ser Pro Met
             20                  25                  30
```

```
Val Gly Thr Phe Tyr Arg Thr Pro Ser Pro Asp Ala Lys Ala Phe Ile
         35                  40                  45

Glu Val Gly Gln Lys Val Asn Val Gly Asp Thr Leu Cys Ile Val Glu
     50                  55                  60

Ala Met Lys Met Met Asn Gln Ile Glu Ala Asp Lys Ser Gly Thr Val
 65                  70                  75                  80

Lys Ala Ile Leu Val Glu Ser Gly Gln Pro Val Glu Phe Asp Glu Pro
                 85                  90                  95

Leu Val Val Ile Glu Gly Ser Glu Leu Glu Ile Gln Met Glu Asp Ala
            100                 105                 110

Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro Leu Glu Asp Gly
        115                 120                 125

Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg Tyr Ala Leu Val
    130                 135                 140

Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu Val Asn Ile Thr
145                 150                 155                 160

Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala Glu Ala Met Lys
                165                 170                 175

Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val Cys Ser Glu Asn
            180                 185                 190

Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu Phe Ile Gly Val
        195                 200                 205

Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg Glu Leu Leu Asn
    210                 215                 220

Ser Met Asn Ile Ser Gln Pro Thr Val Val Phe Val Ser Lys Lys Gly
225                 230                 235                 240

Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro Ile Ile Gln Lys
                245                 250                 255

Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly Phe Gln Ser Met
            260                 265                 270

Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe Asn Glu Tyr Asp
        275                 280                 285

Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile Ala Leu Ile Met
    290                 295                 300

Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val Ala Leu Pro His
305                 310                 315                 320

Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp Pro Ile Phe Gly
                325                 330                 335

Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val Val Pro Phe His
            340                 345                 350

His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu Ile Cys Gly Phe
        355                 360                 365

Arg Val Val Leu Met Tyr Arg Phe Glu Glu Glu Leu Phe Leu Arg Ser
    370                 375                 380

Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val Pro Thr Leu Phe
385                 390                 395                 400

Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr Asp Leu Ser Asn
                405                 410                 415
```

```
Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser Lys Glu Val Gly
            420                 425                 430

Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile Arg Gln Gly Tyr
        435                 440                 445

Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr Pro Glu Gly Asp
    450                 455                 460

Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe Phe Glu Ala Lys
465                 470                 475                 480

Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val Asn Gln Arg Gly
                485                 490                 495

Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly Tyr Val Asn Asn
            500                 505                 510

Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly Trp Leu His Ser
        515                 520                 525

Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe Phe Ile Val Asp
    530                 535                 540

Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln Val Ala Pro Ala
545                 550                 555                 560

Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile Phe Asp Ala Gly
                565                 570                 575

Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu Pro Ala Ala Val
            580                 585                 590

Val Val Leu Glu His Gly Lys Thr Met Thr Glu Lys Glu Ile Val Asp
        595                 600                 605

Tyr Val Ala Ser Gln Val Thr Thr Ala Lys Lys Leu Arg Gly Gly Val
    610                 615                 620

Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly Lys Leu Asp Ala
625                 630                 635                 640

Arg Lys Ile Arg Glu Ile Leu Ile Lys Ala Lys Lys Gly Gly Lys Ser
                645                 650                 655

Lys Leu Ala Ala Ala Met Pro Ala Pro His Gly Gly Ile Leu Gln Asp
            660                 665                 670

Leu Ile Ala Arg Asp Ala Leu Lys Lys Asn Glu Leu Leu Ser Glu Ala
        675                 680                 685

Gln Ser Ser Asp Ile Leu Val Trp Asn Leu Thr Pro Arg Gln Leu Cys
    690                 695                 700

Asp Ile Glu Leu Ile Leu Asn Gly Gly Phe Ser Pro Leu Thr Gly Phe
705                 710                 715                 720

Leu Asn Glu Asn Asp Tyr Ser Ser Val Val Thr Asp Ser Arg Leu Ala
                725                 730                 735

Asp Gly Thr Leu Trp Thr Ile Pro Ile Thr Leu Asp Val Asp Glu Ala
            740                 745                 750

Phe Ala Asn Gln Ile Lys Pro Asp Thr Arg Ile Ala Leu Phe Gln Asp
        755                 760                 765

Asp Glu Ile Pro Ile Ala Ile Leu Thr Val Gln Asp Val Tyr Lys Pro
    770                 775                 780

Asn Lys Thr Ile Glu Ala Glu Lys Val Phe Arg Gly Asp Pro Glu His
```

785 790 795 800

Pro Ala Ile Ser Tyr Leu Phe Asn Val Ala Gly Asp Tyr Tyr Val Gly
805 810 815

Gly Ser Leu Glu Ala Ile Gln Leu Pro Gln His Tyr Asp Tyr Pro Gly
820 825 830

Leu Arg Lys Thr Pro Ala Gln Leu Arg Leu Glu Phe Gln Ser Arg Gln
835 840 845

Trp Asp Arg Val Val Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala
850 855 860

His Arg Glu Leu Thr Val Arg Ala Ala Arg Glu Ala Asn Ala Lys Val
865 870 875 880

Leu Ile His Pro Val Val Gly Leu Thr Lys Pro Gly Asp Ile Asp His
885 890 895

His Thr Arg Val Arg Val Tyr Gln Glu Ile Ile Lys Arg Tyr Pro Asn
900 905 910

Gly Ile Ala Phe Leu Ser Leu Leu Pro Leu Ala Met Arg Met Ser Gly
915 920 925

Asp Arg Glu Ala Val Trp His Ala Ile Ile Arg Lys Asn Tyr Gly Ala
930 935 940

Ser His Phe Ile Val Gly Arg Asp His Ala Gly Pro Gly Lys Asn Ser
945 950 955 960

Lys Gly Val Asp Phe Tyr Gly Pro Tyr Asp Ala Gln Glu Leu Val Glu
965 970 975

Ser Tyr Lys His Glu Leu Asp Ile Glu Val Val Pro Phe Arg Met Val
980 985 990

Thr Tyr Leu Pro Asp Glu Asp Arg Tyr Ala Pro Ile Asp Gln Ile Asp
995 1000 1005

Thr Thr Lys Thr Arg Thr Leu Asn Ile Ser Gly Thr Glu Leu Arg Arg
1010 1015 1020

Arg Leu Arg Val Gly Gly Glu Ile Pro Glu Trp Phe Ser Tyr Pro Glu
1025 1030 1035 1040

Val Val Lys Ile Leu Arg Glu Ser Asn Pro Pro Arg Pro Lys Gln Gly
1045 1050 1055

Phe Ser Ile Val Leu Gly Asn Ser Leu Thr Val Ser Arg Glu Gln Leu
1060 1065 1070

Ser Ile Ala Leu Leu Ser Thr Phe Leu Gln Phe Gly Gly Gly Arg Tyr
1075 1080 1085

Tyr Lys Ile Phe Glu His Asn Asn Lys Thr Glu Leu Leu Ser Leu Ile
1090 1095 1100

Gln Asp Phe Ile Gly Ser Gly Ser Gly Leu Ile Ile Pro Asn Gln Trp
1105 1110 1115 1120

Glu Asp Asp Lys Asp Ser Val Val Gly Lys Gln Asn Val Tyr Leu Leu
1125 1130 1135

Asp Thr Ser Ser Ser Ala Asp Ile Gln Leu Glu Ser Ala Asp Glu Pro
1140 1145 1150

Ile Ser His Ile Val Gln Lys Val Val Leu Phe Leu Glu Asp Asn Gly
1155 1160 1165

Phe Phe Val Phe
1170

<210> 5
<211> 1488
<212> DNA
<213> Escherichia coli

<400> 5
atgcggggtt ctcatcatca tcatcatcat ggtatggcta gcatggaagc gccagcagca 60
gcggaaatca gtggtcacat cgtacgttcc ccgatggttg gtactttcta ccgcacccca 120
agcccggacg caaaagcgtt catcgaagtg ggtcagaaag tcaacgtggg cgataccctg 180
tgcatcgttg aagccatgaa aatgatgaac cagatcgaag cggacaaatc cggtaccgtg 240
aaagcaattc tggtcgaaag tggacaaccg gtagaatttg acgagccgct ggtcgtcatc 300
gagggatccg agctcgagat ctgcagcatg agcgtaagca tcccgcatgg cggcacattg 360
atcaaccgtt ggaatccgga ttacccaatc gatgaagcaa cgaaaacgat cgagctgtcc 420
aaagccgaac taagcgacct tgagctgatc ggcacaggcg cctacagccc gctcaccggg 480
tttttaacga aagccgatta cgatgcggtc gtagaaacga tgcgcctcgc tgatggcact 540
gtctggagca ttccgatcac gctggcggtg acggaagaaa aagcgagtga actcactgtc 600
ggcgacaaag cgaaactcgt ttatggcggc gacgtctacg gcgtcattga aatcgccgat 660
atttaccgcc cggataaaac gaaagaagcc aagctcgtct ataaaaccga tgaactcgct 720
cacccgggcg tgcgcaagct gtttgaaaaa ccagatgtgt acgtcggcgg agcggttacg 780
ctcgtcaaac ggaccgacaa aggccagttt gctccgtttt atttcgatcc ggccgaaacg 840
cggaaacgat ttgccgaact cggctggaat accgtcgtcg gcttccaaac acgcaacccg 900
gttcaccgcg cccatgaata cattcaaaaa tgcgcgcttg aaatcgtgga cggcttgttt 960
ttaaacccgc tcgtcggcga aacgaaagcg gacgatattc cggccgacat ccggatggaa 1020
agctatcaag tgctgctgga aaactattat ccgaaagacc gcgttttctt gggcgtcttc 1080
caagctgcga tgcgctatgc cggtccgcgc gaagcgattt tccatgccat ggtgcggaaa 1140
aacttcggct gcacgcactt catcgtcggc cgcgaccatg cgggcgtcgg caactattac 1200
ggcacgtatg atgcgcaaaa aatcttctcg aactttacag ccgaagagct tggcattaca 1260
ccgctctttt tcgaacacag cttttattgc acgaaatgcg aaggcatggc atcgacgaaa 1320
acatgcccgc acgacgcaca atatcacgtt gtcctttctg gcacgaaagt ccgtgaaatg 1380
ttgcgtaacg gccaagtgcc gccgagcaca ttcagccgtc cggaagtggc cgccgttttg 1440
atcaaagggc tgcaagaacg cgaaacggtc gccccgtcag cgcgctaa 1488

<210> 6
<211> 495
<212> PRT
<213> Escherichia coli

<400> 6
Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Glu
1 5 10 15

Ala Pro Ala Ala Ala Glu Ile Ser Gly His Ile Val Arg Ser Pro Met
20 25 30

Val Gly Thr Phe Tyr Arg Thr Pro Ser Pro Asp Ala Lys Ala Phe Ile
35 40 45

Glu Val Gly Gln Lys Val Asn Val Gly Asp Thr Leu Cys Ile Val Glu
50 55 60

Ala Met Lys Met Met Asn Gln Ile Glu Ala Asp Lys Ser Gly Thr Val
65 70 75 80

Lys Ala Ile Leu Val Glu Ser Gly Gln Pro Val Glu Phe Asp Glu Pro
85 90 95

Leu Val Val Ile Glu Gly Ser Glu Leu Glu Ile Cys Ser Met Ser Val
100 105 110

Ser Ile Pro His Gly Gly Thr Leu Ile Asn Arg Trp Asn Pro Asp Tyr
115 120 125

Pro Ile Asp Glu Ala Thr Lys Thr Ile Glu Leu Ser Lys Ala Glu Leu
130 135 140

Ser Asp Leu Glu Leu Ile Gly Thr Gly Ala Tyr Ser Pro Leu Thr Gly

145 150 155 160

Phe Leu Thr Lys Ala Asp Tyr Asp Ala Val Val Glu Thr Met Arg Leu
165 170 175

Ala Asp Gly Thr Val Trp Ser Ile Pro Ile Thr Leu Ala Val Thr Glu
180 185 190

Glu Lys Ala Ser Glu Leu Thr Val Gly Asp Lys Ala Lys Leu Val Tyr
195 200 205

Gly Gly Asp Val Tyr Gly Val Ile Glu Ile Ala Asp Ile Tyr Arg Pro
210 215 220

Asp Lys Thr Lys Glu Ala Lys Leu Val Tyr Lys Thr Asp Glu Leu Ala
225 230 235 240

His Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val Tyr Val Gly
245 250 255

Gly Ala Val Thr Leu Val Lys Arg Thr Asp Lys Gly Gln Phe Ala Pro
260 265 270

Phe Tyr Phe Asp Pro Ala Glu Thr Arg Lys Arg Phe Ala Glu Leu Gly
275 280 285

Trp Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val His Arg Ala
290 295 300

His Glu Tyr Ile Gln Lys Cys Ala Leu Glu Ile Val Asp Gly Leu Phe
305 310 315 320

Leu Asn Pro Leu Val Gly Glu Thr Lys Ala Asp Asp Ile Pro Ala Asp
325 330 335

Ile Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr Tyr Pro Lys
340 345 350

Asp Arg Val Phe Leu Gly Val Phe Gln Ala Ala Met Arg Tyr Ala Gly
355 360 365

Pro Arg Glu Ala Ile Phe His Ala Met Val Arg Lys Asn Phe Gly Cys
370 375 380

Thr His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly Asn Tyr Tyr
385 390 395 400

Gly Thr Tyr Asp Ala Gln Lys Ile Phe Ser Asn Phe Thr Ala Glu Glu
405 410 415

Leu Gly Ile Thr Pro Leu Phe Phe Glu His Ser Phe Tyr Cys Thr Lys
420 425 430

Cys Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp Ala Gln Tyr
435 440 445

His Val Val Leu Ser Gly Thr Lys Val Arg Glu Met Leu Arg Asn Gly
450 455 460

Gln Val Pro Pro Ser Thr Phe Ser Arg Pro Glu Val Ala Ala Val Leu
465 470 475 480

Ile Lys Gly Leu Gln Glu Arg Glu Thr Val Ala Pro Ser Ala Arg
485 490 495

<210> 7
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 7
cccttctgca gcatgagcgt aagcatcccg catggcggca cattg 45

<210> 8
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 8
cccgtaagct tttagcgcgc tgacggggcg accgtttcgc gttcttg 47

<210> 9
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 9
ccccctcgag atccaaatgg aagacgccaa aaacataaag aaaggccc 48

<210> 10
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 10
ccccctcgag atccaaatgg ctgacaaaaa catcctgtat ggccc 45

<210> 11
<211> 67
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 11
ttgtagaata ccaccgtgag gagcaggcat agcggccgcc aatttggact ttccgccctt 60
cttggcc 67

<210> 12
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 12
ttgtagaata ccaccgtgag gagcaggcat agcggccgca ccgttggtgt gtttctcgaa 60
catc 64

<210> 13
<211> 37

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 13
gcggccgcta tgcctgctcc tcacggtggt attctac 37

<210> 14
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 14
ccccaagctt ttaaaataca aaaaagccat tgtcttccaa gaataggac 49

<210> 15
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 15
ccccggatcc atccaaatgc ctgctcctca cggtggtatt ctacaagac 49

<210> 16
<211> 62
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 16
gggcctttct ttatgttttt ggcgtcttcc atagcggccg caaatacaaa aaagccattg 60
tc 62

<210> 17
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 17
gcggccgcta tggaagacgc caaaaacata aagaaaggcc c 41

<210> 18
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 18
ccccccatgg ttacaatttg gactttccgc ccttcttggc c 41

<210> 19
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 19
gggccataca ggatgttttt gtcagccata gcggccgcaa atacaaaaaa gccattgtc 59

<210> 20
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 20
gcggccgcta tggctgacaa aaacatcctg tatggccc 38

<210> 21
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 21
ccccaagctt ctaaccgttg gtgtgtttct cgaacatctg acgc 44

<210> 22
<211> 386
<212> PRT
<213> Bacillus stearothermophilus

<400> 22
Met Ser Val Ser Ile Pro His Gly Gly Thr Leu Ile Asn Arg Trp Asn
1 5 10 15

Pro Asp Tyr Pro Leu Asp Glu Ala Thr Lys Thr Ile Glu Leu Ser Lys
20 25 30

Ala Glu Leu Ser Asp Leu Glu Leu Ile Gly Thr Gly Ala Tyr Ser Pro
35 40 45

Leu Thr Gly Phe Leu Thr Lys Thr Asp Tyr Asp Ala Val Val Glu Thr
50 55 60

Met Arg Leu Ser Asp Gly Thr Val Trp Ser Ile Pro Val Thr Leu Ala
65 70 75 80

Val Thr Glu Glu Lys Ala Lys Glu Leu Ala Val Gly Asp Lys Ala Lys
85 90 95

Leu Val Tyr Arg Gly Asp Val Tyr Gly Val Ile Glu Ile Ala Asp Ile
100 105 110

Tyr Arg Pro Asp Lys Thr Lys Glu Ala Lys Leu Val Tyr Lys Thr Asp
115 120 125

Glu Leu Ala His Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val
130 135 140

Tyr Val Gly Gly Glu Ile Thr Leu Val Lys Arg Thr Asp Lys Gly Gln
145 150 155 160

```
Phe Ala Ala Phe Tyr Phe Asp Pro Ala Glu Thr Arg Lys Lys Phe Ala
                165                 170                 175

Glu Phe Gly Trp Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val
            180                 185                 190

His Arg Ala His Glu Tyr Ile Gln Lys Cys Ala Leu Glu Ile Val Asp
        195                 200                 205

Gly Leu Phe Leu Asn Pro Leu Val Gly Glu Thr Lys Ser Asp Asp Ile
    210                 215                 220

Pro Ala Asp Ile Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr
225                 230                 235                 240

Tyr Pro Lys Asp Arg Val Phe Leu Gly Val Phe Gln Ala Ala Met Arg
                245                 250                 255

Tyr Ala Gly Pro Arg Glu Ala Ile Phe His Ala Met Val Arg Lys Asn
            260                 265                 270

Phe Gly Cys Thr His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly
        275                 280                 285

Asn Tyr Tyr Gly Thr Tyr Asp Ala Gln Lys Ile Phe Leu Asn Phe Thr
    290                 295                 300

Ala Glu Glu Leu Gly Ile Thr Pro Leu Phe Phe Glu His Ser Phe Tyr
305                 310                 315                 320

Cys Thr Lys Cys Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp
                325                 330                 335

Ala Lys Tyr His Val Val Leu Ser Gly Thr Lys Val Arg Glu Met Leu
            340                 345                 350

Arg Asn Gly Gln Val Pro Pro Ser Thr Phe Ser Arg Pro Glu Val Ala
        355                 360                 365

Ala Val Leu Ile Lys Gly Leu Gln Glu Arg Glu Thr Val Ala Pro Ser
    370                 375                 380

Ala Arg
385


<210> 23
<211> 546
<212> PRT
<213> Aquifex aeolicus

<400> 23
Met Glu Lys Ile Lys Tyr Leu Lys Ser Ile Gln Ile Ser Gln Arg Ser
  1               5                  10                  15

Val Leu Asp Leu Lys Leu Leu Ala Val Gly Ala Phe Thr Pro Leu Asp
             20                  25                  30

Arg Phe Met Gly Glu Glu Asp Tyr Arg Asn Val Val Glu Ser Met Arg
         35                  40                  45

Leu Lys Ser Gly Thr Leu Phe Pro Ile Pro Ile Thr Leu Pro Met Glu
     50                  55                  60

Lys Glu Ile Ala Lys Asp Leu Lys Glu Gly Glu Trp Ile Val Leu Arg
 65                  70                  75                  80

Asp Pro Lys Asn Val Pro Leu Ala Ile Met Arg Val Glu Glu Val Tyr
                 85                  90                  95
```

```
Lys Trp Asn Leu Glu Tyr Glu Ala Lys Asn Val Leu Gly Thr Thr Asp
            100                 105                 110

Pro Arg His Pro Leu Val Ala Glu Met His Thr Trp Gly Glu Tyr Tyr
        115                 120                 125

Ile Ser Gly Glu Leu Lys Val Ile Gln Leu Pro Lys Tyr Tyr Asp Phe
    130                 135                 140

Pro Glu Tyr Arg Lys Thr Pro Lys Gln Val Arg Glu Glu Ile Lys Ser
145                 150                 155                 160

Leu Gly Leu Asp Lys Ile Val Ala Phe Gln Thr Arg Asn Pro Met His
                165                 170                 175

Arg Val His Glu Glu Leu Thr Lys Arg Ala Met Glu Lys Val Gly Gly
            180                 185                 190

Gly Leu Leu Leu His Pro Val Val Gly Leu Thr Lys Pro Gly Asp Val
        195                 200                 205

Asp Val Tyr Thr Arg Met Arg Ile Tyr Lys Val Leu Tyr Glu Lys Tyr
    210                 215                 220

Tyr Asp Lys Lys Lys Thr Ile Leu Ala Phe Leu Pro Leu Ala Met Arg
225                 230                 235                 240

Met Ala Gly Pro Arg Glu Ala Leu Trp His Gly Ile Ile Arg Arg Asn
                245                 250                 255

Tyr Gly Ala Thr His Phe Ile Val Gly Arg Asp His Ala Ser Pro Gly
            260                 265                 270

Lys Asp Ser Lys Gly Lys Pro Phe Tyr Asp Pro Tyr Glu Ala Gln Glu
        275                 280                 285

Leu Phe Lys Lys Tyr Glu Asp Glu Ile Gly Ile Lys Met Val Pro Phe
    290                 295                 300

Glu Glu Leu Val Tyr Val Pro Glu Leu Asp Gln Tyr Val Glu Ile Asn
305                 310                 315                 320

Glu Ala Lys Lys Arg Asn Leu Lys Tyr Ile Asn Ile Ser Gly Thr Glu
                325                 330                 335

Ile Arg Glu Asn Phe Leu Lys Gln Gly Arg Lys Leu Pro Glu Trp Phe
            340                 345                 350

Thr Arg Pro Glu Val Ala Glu Ile Leu Ala Glu Thr Tyr Val Pro Lys
        355                 360                 365

His Lys Gln Gly Phe Cys Val Trp Leu Thr Gly Leu Pro Cys Ala Gly
    370                 375                 380

Lys Ser Thr Ile Ala Glu Ile Leu Ala Thr Met Leu Gln Ala Arg Gly
385                 390                 395                 400

Arg Lys Val Thr Leu Leu Asp Gly Asp Val Val Arg Thr His Leu Ser
                405                 410                 415

Arg Gly Leu Gly Phe Ser Lys Glu Asp Arg Ile Thr Asn Ile Leu Arg
            420                 425                 430

Val Gly Phe Val Ala Ser Glu Ile Val Lys His Asn Gly Val Val Ile
        435                 440                 445

Cys Ala Leu Val Ser Pro Tyr Arg Ser Ala Arg Asn Gln Val Arg Asn
    450                 455                 460

Met Met Glu Glu Gly Lys Phe Ile Glu Val Phe Val Asp Ala Pro Val
465                 470                 475                 480
```

```
Glu Val Cys Glu Glu Arg Asp Val Lys Gly Leu Tyr Lys Lys Ala Lys
                485                 490                 495

Glu Gly Leu Ile Lys Gly Phe Thr Gly Val Asp Asp Pro Tyr Glu Pro
            500                 505                 510

Pro Val Ala Pro Glu Val Arg Val Asp Thr Thr Lys Leu Thr Pro Glu
        515                 520                 525

Glu Ser Ala Leu Lys Ile Leu Glu Phe Leu Lys Lys Glu Gly Phe Ile
    530                 535                 540

Lys Asp
545
```

<210> 24
<211> 259
<212> PRT
<213> Pyrococcus furiosus

```
<400> 24
Met Val Ser Lys Pro His Gly Gly Lys Leu Ile Arg Arg Ile Ala Ala
  1               5                  10                  15

Pro Arg Thr Arg Glu Arg Ile Leu Ser Glu Gln His Glu Tyr Pro Arg
             20                  25                  30

Val Gln Ile Asp His Gly Arg Ala Ile Asp Leu Glu Asn Ile Ala His
         35                  40                  45

Gly Val Tyr Ser Pro Leu Lys Gly Phe Leu Thr Arg Glu Asp Phe Glu
     50                  55                  60

Ser Val Leu Asp Tyr Met Arg Leu Ser Asp Asp Thr Pro Trp Thr Ile
 65                  70                  75                  80

Pro Ile Val Leu Asp Val Gly Glu Pro Thr Phe Glu Gly Gly Asp Ala
                 85                  90                  95

Ile Leu Leu Tyr Tyr Glu Asn Pro Pro Ile Ala Arg Met His Val Glu
            100                 105                 110

Asp Ile Tyr Thr Tyr Asp Lys Lys Glu Phe Ala Val Lys Val Phe Lys
        115                 120                 125

Thr Asp Asp Pro Asn His Leu Gly Val Ala Arg Val Tyr Ser Met Gly
    130                 135                 140

Lys Tyr Leu Val Gly Gly Gly Ile Glu Leu Leu Asn Glu Leu Pro Asn
145                 150                 155                 160

Pro Phe Ala Lys Tyr Thr Leu Arg Pro Val Glu Thr Arg Ile Leu Phe
                165                 170                 175

Lys Glu Arg Gly Trp Lys Thr Ile Val Ala Phe Gln Thr Arg Asn Val
            180                 185                 190

Pro His Leu Gly His Glu Tyr Val Gln Lys Ala Ala Leu Thr Phe Val
        195                 200                 205

Asp Gly Leu Phe Ile Asn Pro Val Leu Gly Arg Lys Lys Lys Gly Asp
    210                 215                 220

Tyr Lys Asp Glu Val Ile Ile Lys Ala Tyr Tyr Leu Ile Met Lys Tyr
225                 230                 235                 240

Cys Ser Asn Thr Thr His His Ala Ile Met Arg Lys Thr Ser Thr Ser
                245                 250                 255
```

Ser Gln Thr

<210> 25
<211> 406
<212> PRT
<213> Sulfolobus solfataricus

<400> 25
Met Asn Leu Ile Gly His Gly Lys Val Glu Ile Val Glu Arg Ile Lys
1 5 10 15

Thr Ile Ser Asp Phe Lys Glu Leu His Arg Ile Glu Val Lys Arg Gln
20 25 30

Leu Ala His Glu Ile Val Ser Ile Ala Tyr Gly Phe Leu Ser Pro Leu
35 40 45

Lys Gly Phe Met Asn Tyr Glu Glu Val Asp Gly Val Val Glu Asn Met
50 55 60

Arg Leu Pro Asn Gly Val Leu Trp Pro Ile Pro Leu Val Phe Asp Tyr
65 70 75 80

Ser Gln Asn Glu Lys Val Lys Glu Gly Asp Thr Ile Gly Ile Thr Tyr
85 90 95

Leu Gly Lys Pro Leu Ala Ile Met Lys Val Lys Glu Ile Phe Lys Tyr
100 105 110

Asp Lys Leu Lys Ile Ala Glu Lys Val Tyr Lys Thr Lys Asp Ile Lys
115 120 125

His Pro Gly Val Lys Arg Thr Leu Ser Tyr Ala Asp Ala Phe Leu Ala
130 135 140

Gly Asp Val Trp Leu Val Arg Glu Pro Gln Phe Asn Lys Pro Tyr Ser
145 150 155 160

Glu Phe Trp Leu Thr Pro Arg Met His Arg Thr Val Phe Glu Lys Lys
165 170 175

Gly Trp Lys Arg Val Val Ala Phe Gln Thr Arg Asn Val Pro His Thr
180 185 190

Gly His Glu Tyr Leu Met Lys Phe Ala Trp Phe Ala Ala Asn Glu Asn
195 200 205

Gln Lys Val Asp Glu Pro Arg Thr Gly Ile Leu Val Asn Val Val Ile
210 215 220

Gly Glu Lys Arg Val Gly Asp Tyr Ile Asp Glu Ala Ile Leu Leu Thr
225 230 235 240

His Asp Ala Leu Ser Lys Tyr Gly Tyr Ile Ser Pro Lys Val His Leu
245 250 255

Leu Ser Phe Thr Leu Trp Asp Met Arg Tyr Ala Gly Pro Arg Glu Ala
260 265 270

Leu Leu His Ala Ile Ile Arg Ser Asn Leu Gly Cys Thr His His Val
275 280 285

Phe Gly Arg Asp His Ala Gly Val Gly Asn Tyr Tyr Ser Pro Tyr Glu
290 295 300

Ala His Glu Ile Phe Asp Ser Ile Asn Glu Glu Asp Leu Leu Ile Lys
305 310 315 320

Pro Ile Phe Leu Arg Glu Asn Tyr Tyr Cys Pro Arg Cys Gly Ser Ile

325 330 335

Glu Asn Glu Ile Leu Cys Asp His Lys Asp Glu Lys Gln Glu Phe Ser
340 345 350

Gly Ser Leu Ile Arg Ser Ile Ile Leu Asp Glu Val Lys Pro Thr Lys
355 360 365

Met Val Met Arg Pro Glu Val Tyr Asp Val Leu Met Lys Ala Ala Glu
370 375 380

Gln Tyr Gly Phe Gly Ser Pro Phe Val Thr Glu Glu Tyr Leu Glu Lys
385 390 395 400

Arg Gln Ser Ile Leu Gly
405

<210> 26
<211> 455
<212> PRT
<213> Pyrobaculum aerophilum

<400> 26
Met Pro Met Pro Ala Pro Leu Glu Pro His Gly Gly Arg Leu Val Tyr
1 5 10 15

Asn Val Ile Glu Asp Arg Asp Lys Ala Ala Ala Met Ile Gln Gly Leu
20 25 30

Pro Ser Ile Glu Ile Glu Pro Thr Leu Gly Pro Asp Gly Ser Pro Ile
35 40 45

Arg Asn Pro Tyr Arg Glu Ile Met Ser Ile Ala Tyr Gly Phe Phe Ser
50 55 60

Pro Val Glu Gly Phe Met Thr Arg Asn Glu Val Glu Ser Ile Leu Lys
65 70 75 80

Glu Arg Arg Leu Leu Asn Gly Trp Leu Phe Pro Phe Pro Leu Ile Tyr
85 90 95

Asp Val Asp Glu Glu Lys Ile Lys Gly Ile Lys Glu Gly Asp Ser Val
100 105 110

Leu Leu Lys Leu Lys Gly Lys Pro Leu Ala Val Leu Asn Val Glu Glu
115 120 125

Ile Trp Arg Leu Pro Asp Arg Lys Glu Leu Ala Asp Ala Val Phe Gly
130 135 140

Thr Pro Glu Arg Asn Lys Glu Val Val Lys Lys Arg Phe Asp Glu Lys
145 150 155 160

His Pro Gly Trp Leu Ile Tyr Arg Ser Met Arg Pro Met Ala Leu Ala
165 170 175

Gly Lys Ile Thr Val Val Asn Pro Pro Arg Phe Lys Glu Pro Tyr Ser
180 185 190

Arg Phe Trp Met Pro Pro Arg Val Ser Arg Glu Tyr Val Glu Lys Lys
195 200 205

Gly Trp Arg Ile Val Val Ala His Gln Thr Arg Asn Val Pro His Ile
210 215 220

Gly His Glu Met Leu Met Lys Arg Ala Met Phe Val Ala Gly Gly Glu
225 230 235 240

Arg Pro Gly Asp Ala Val Leu Val Asn Ala Ile Ile Gly Ala Lys Arg
245 250 255

```
Pro Gly Asp Tyr Val Asp Glu Ala Ile Leu Glu Gly His Glu Ala Leu
            260                 265                 270

Asn Lys Ala Gly Tyr Phe His Pro Asp Arg His Val Val Thr Met Thr
        275                 280                 285

Leu Trp Asp Met Arg Tyr Gly Asn Pro Leu Glu Ser Leu Leu His Gly
    290                 295                 300

Ile Ile Arg Gln Asn Met Gly Ala Thr His His Met Phe Gly Arg Asp
305                 310                 315                 320

His Ala Ala Thr Gly Asp Tyr Tyr Asp Pro Tyr Ala Thr Gln Tyr Leu
                325                 330                 335

Trp Thr Arg Gly Leu Pro Ser Tyr Gly Leu Asn Glu Pro Pro His Met
            340                 345                 350

Thr Asp Lys Gly Leu Arg Ile Lys Pro Val Asn Leu Gly Glu Phe Ala
        355                 360                 365

Tyr Cys Pro Lys Cys Gly Glu Tyr Thr Tyr Leu Gly Met Ser Tyr Glu
    370                 375                 380

Gly Tyr Lys Glu Val Ala Leu Cys Gly His Thr Pro Glu Arg Ile Ser
385                 390                 395                 400

Gly Ser Leu Leu Arg Gly Ile Ile Ile Glu Gly Leu Arg Pro Pro Lys
                405                 410                 415

Val Val Met Arg Pro Glu Val Tyr Asp Val Ile Val Lys Trp Trp Arg
            420                 425                 430

Val Tyr Gly Tyr Pro Tyr Val Thr Asp Lys Tyr Leu Arg Ile Lys Glu
        435                 440                 445

Gln Glu Leu Glu Val Glu Leu
    450                 455
```

<210> 27
<211> 456
<212> PRT
<213> Archaeoglobus fulgidus

<400> 27

```
Met Pro Leu Ile Lys Thr Pro Pro Pro His Gly Gly Lys Leu Val Glu
  1               5                  10                  15

Arg Val Val Lys Lys Arg Asp Ile Ala Glu Lys Met Ile Ala Gly Cys
            20                  25                  30

Pro Thr Tyr Glu Leu Lys Pro Thr Thr Leu Pro Asp Gly Thr Pro Ile
        35                  40                  45

Arg His Val Tyr Arg Glu Ile Met Ser Val Cys Tyr Gly Phe Phe Ser
    50                  55                  60

Pro Val Glu Gly Ser Met Val Gln Asn Glu Leu Glu Arg Val Leu Asn
65                  70                  75                  80

Glu Arg Arg Leu Leu Ser Glu Trp Ile Phe Pro Tyr Pro Ile Leu Phe
                85                  90                  95

Asp Ile Ser Glu Glu Asp Tyr Lys Ala Leu Asp Val Lys Glu Gly Asp
            100                 105                 110

Arg Leu Leu Leu Met Leu Lys Gly Gln Pro Phe Ala Thr Leu Asp Ile
        115                 120                 125
```

```
Glu Glu Val Tyr Lys Ile Asp Pro Val Asp Val Ala Thr Arg Thr Phe
    130                 135                 140

Gly Thr Pro Glu Lys Asn Pro Glu Val Val Arg Glu Pro Phe Asp Asp
145                 150                 155                 160

Lys His Pro Gly Tyr Val Ile Tyr Lys Met His Asn Pro Ile Ile Leu
                165                 170                 175

Ala Gly Lys Tyr Thr Ile Val Asn Glu Pro Lys Phe Lys Glu Pro Tyr
            180                 185                 190

Asp Arg Phe Trp Phe Pro Pro Ser Lys Cys Arg Glu Val Ile Lys Asn
        195                 200                 205

Glu Lys Lys Trp Arg Thr Val Ile Ala His Gln Thr Arg Asn Val Pro
    210                 215                 220

His Val Gly His Glu Met Leu Met Lys Cys Ala Ala Tyr Thr Gly Asp
225                 230                 235                 240

Ile Glu Pro Cys His Gly Ile Leu Val Asn Ala Ile Ile Gly Ala Lys
                245                 250                 255

Arg Arg Gly Asp Tyr Pro Asp Glu Ala Ile Leu Glu Gly His Glu Ala
            260                 265                 270

Val Asn Lys Tyr Gly Tyr Ile Lys Pro Glu Arg His Met Val Thr Phe
        275                 280                 285

Thr Leu Trp Asp Met Arg Tyr Gly Asn Pro Ile Glu Ser Leu Leu His
    290                 295                 300

Gly Val Ile Arg Gln Asn Met Gly Cys Thr His His Met Phe Gly Arg
305                 310                 315                 320

Asp His Ala Ala Val Gly Glu Tyr Tyr Asp Met Tyr Ala Thr Gln Ile
                325                 330                 335

Leu Trp Ser Gln Gly Ile Pro Ser Phe Gly Phe Glu Ala Pro Pro Asn
            340                 345                 350

Glu Val Asp Tyr Gly Leu Lys Ile Ile Pro Gln Asn Met Ala Glu Phe
        355                 360                 365

Trp Tyr Cys Pro Ile Cys Gln Glu Ile Ala Tyr Ser Glu Asn Cys Gly
    370                 375                 380

His Thr Asp Ala Lys Gln Lys Phe Ser Gly Ser Phe Leu Arg Gly Met
385                 390                 395                 400

Val Ala Glu Gly Val Phe Pro Pro Arg Val Val Met Arg Pro Glu Val
                405                 410                 415

Tyr Lys Gln Ile Val Lys Trp Trp Lys Val Tyr Asn Tyr Pro Phe Val
            420                 425                 430

Asn Arg Lys Tyr Leu Glu Leu Lys Asn Lys Glu Leu Glu Ile Asp Leu
        435                 440                 445

Pro Ala Met Glu Val Pro Lys Ala
    450                 455
```

<210> 28
<211> 573
<212> PRT
<213> Penicillium chrysogenum

<400> 28
Met Ala Asn Ala Pro His Gly Gly Val Leu Lys Asp Leu Leu Ala Arg

```
  1               5                  10                  15
Asp Ala Pro Arg Gln Ala Glu Leu Ala Ala Glu Ala Glu Ser Leu Pro
            20                  25                  30
Ala Val Thr Leu Thr Glu Arg Gln Leu Cys Asp Leu Glu Leu Ile Met
        35                  40                  45
Asn Gly Gly Phe Ser Pro Leu Glu Gly Phe Met Asn Gln Ala Asp Tyr
    50                  55                  60
Asp Arg Val Cys Glu Asp Asn Arg Leu Ala Asp Gly Asn Val Phe Ser
65                  70                  75                  80
Met Pro Ile Thr Leu Asp Ala Ser Gln Glu Val Ile Asp Glu Lys Lys
                85                  90                  95
Leu Gln Ala Ala Ser Arg Ile Thr Leu Arg Asp Phe Arg Asp Asp Arg
            100                 105                 110
Asn Leu Ala Ile Leu Thr Ile Asp Asp Ile Tyr Arg Pro Asp Lys Thr
        115                 120                 125
Lys Glu Ala Lys Leu Val Phe Gly Gly Asp Pro Glu His Pro Ala Ile
    130                 135                 140
Val Tyr Leu Asn Asn Thr Val Lys Glu Phe Tyr Ile Gly Gly Lys Ile
145                 150                 155                 160
Glu Ala Val Asn Lys Leu Asn His Tyr Asp Tyr Val Ala Leu Arg Tyr
                165                 170                 175
Thr Pro Ala Glu Leu Arg Val His Phe Asp Lys Leu Gly Trp Ser Arg
            180                 185                 190
Val Val Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala His Arg Glu
        195                 200                 205
Leu Thr Val Arg Ala Ala Arg Ser Arg Gln Ala Asn Val Leu Ile His
    210                 215                 220
Pro Val Val Gly Leu Thr Lys Pro Gly Asp Ile Asp His Phe Thr Arg
225                 230                 235                 240
Val Arg Ala Tyr Gln Ala Leu Leu Pro Arg Tyr Pro Asn Gly Met Ala
                245                 250                 255
Val Leu Gly Leu Leu Gly Leu Ala Met Arg Met Gly Gly Pro Arg Glu
            260                 265                 270
Ala Ile Trp His Ala Ile Ile Arg Lys Asn His Gly Ala Thr His Phe
        275                 280                 285
Ile Val Gly Arg Asp His Ala Gly Pro Gly Ser Asn Ser Lys Gly Glu
    290                 295                 300
Asp Phe Tyr Gly Pro Tyr Asp Ala Gln His Ala Val Glu Lys Tyr Lys
305                 310                 315                 320
Asp Glu Leu Gly Ile Glu Val Val Glu Phe Gln Met Val Thr Tyr Leu
                325                 330                 335
Pro Asp Thr Asp Glu Tyr Arg Pro Val Asp Gln Val Pro Ala Gly Val
            340                 345                 350
Lys Thr Leu Asn Ile Ser Gly Thr Glu Leu Arg Arg Arg Leu Arg Ser
        355                 360                 365
Gly Ala His Ile Pro Glu Trp Phe Ser Tyr Pro Glu Val Val Lys Ile
    370                 375                 380
```

```
Leu Arg Glu Ser Asn Pro Pro Arg Ala Thr Gln Gly Phe Thr Ile Phe
385                 390                 395                 400

Leu Thr Gly Tyr Met Asn Ser Gly Lys Asp Ala Ile Ala Arg Ala Leu
                405                 410                 415

Gln Val Thr Leu Asn Gln Gln Gly Gly Arg Ser Val Ser Leu Leu Leu
            420                 425                 430

Gly Asp Thr Val Arg His Glu Leu Ser Ser Glu Leu Gly Phe Thr Arg
        435                 440                 445

Glu Asp Arg His Thr Asn Ile Gln Arg Ile Ala Phe Val Ala Thr Glu
    450                 455                 460

Leu Thr Arg Ala Gly Ala Ala Val Ile Ala Ala Pro Ile Ala Pro Tyr
465                 470                 475                 480

Glu Glu Ser Arg Lys Phe Ala Arg Asp Ala Val Ser Gln Ala Gly Ser
                485                 490                 495

Phe Phe Leu Val His Val Ala Thr Pro Leu Glu His Cys Glu Gln Ser
            500                 505                 510

Asp Lys Arg Gly Ile Tyr Ala Ala Ala Arg Arg Gly Glu Ile Lys Gly
        515                 520                 525

Phe Thr Gly Val Asp Asp Pro Tyr Glu Thr Pro Glu Lys Ala Asp Leu
    530                 535                 540

Val Val Asp Phe Ser Lys Gln Ser Val Arg Ser Ile Val His Glu Ile
545                 550                 555                 560

Ile Leu Val Leu Glu Ser Gln Gly Phe Leu Glu Arg Gln
                565                 570
```

<210> 29
<211> 389
<212> PRT
<213> Aeropyrum pernix

<400> 29

```
Met Gly Cys Ser Val Gly Leu Val Ser Arg Pro His Gly Gly Arg Leu
  1               5                  10                  15

Val Arg Arg Val Leu Ser Gly Arg Arg Arg Glu Ile Phe Glu Ser Gln
             20                  25                  30

Tyr Arg Glu Met Pro Arg Leu Glu Val Pro Leu Glu Arg Ala Ile Asp
         35                  40                  45

Ala Glu Asp Leu Ala Arg Gly Val Phe Ser Pro Leu Glu Gly Phe Met
     50                  55                  60

Val Glu Asp Asp Tyr Leu Ser Val Leu Ser Arg Met Arg Leu Ser Asn
 65                  70                  75                  80

Asp Leu Pro Trp Thr Ile Pro Ile Val Leu Asp Ala Asn Arg Glu Trp
                 85                  90                  95

Val Leu Asn Glu Gly Val Ser Ala Gly Asp Asp Ile Ile Leu Thr Tyr
            100                 105                 110

His Gly Leu Pro Ile Ala Val Leu Thr Leu Glu Asp Ile Tyr Ser Trp
        115                 120                 125

Asp Lys Gly Leu His Ala Glu Lys Val Phe Lys Thr Arg Asp Pro Asn
    130                 135                 140

His Pro Gly Val Glu Ala Thr Tyr Lys Arg Gly Asp Ile Leu Leu Gly
```

145 150 155 160

Gly Arg Leu Glu Leu Ile Gln Gly Pro Pro Asn Pro Leu Glu Arg Tyr
165 170 175

Thr Leu Trp Pro Val Glu Thr Arg Val Leu Phe Lys Glu Lys Gly Trp
180 185 190

Arg Thr Val Ala Ala Phe Gln Thr Arg Asn Val Pro His Leu Gly His
195 200 205

Glu Tyr Val Gln Lys Ala Ala Leu Thr Phe Val Asp Gly Leu Leu Val
210 215 220

His Pro Leu Ala Gly Trp Lys Lys Arg Gly Asp Tyr Arg Asp Glu Val
225 230 235 240

Ile Ile Arg Ala Tyr Glu Ala Leu Ile Thr His Tyr Tyr Pro Arg Gly
245 250 255

Val Val Val Leu Ser Val Leu Arg Met Asn Met Asn Tyr Ala Gly Pro
260 265 270

Arg Glu Ala Val His His Ala Ile Val Arg Lys Asn Phe Gly Ala Thr
275 280 285

His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly Ser Tyr Tyr Gly
290 295 300

Pro Tyr Glu Ala Trp Glu Ile Phe Arg Glu Phe Pro Asp Leu Gly Ile
305 310 315 320

Thr Pro Leu Phe Val Arg Glu Ala Tyr Tyr Cys Arg Arg Cys Gly Gly
325 330 335

Met Val Asn Glu Lys Val Cys Pro His Gly Asp Glu Tyr Arg Val Arg
340 345 350

Ile Ser Gly Thr Arg Leu Arg Glu Met Leu Gly Arg Gly Glu Arg Pro
355 360 365

Pro Glu Tyr Met Met Arg Pro Glu Val Ala Asp Ala Ile Ile Ser His
370 375 380

Pro Asp Pro Phe Ile
385

<210> 30
<211> 511
<212> PRT
<213> Saccharomyces cerevisiae

<400> 30
Met Pro Ala Pro His Gly Gly Ile Leu Gln Asp Leu Ile Ala Arg Asp
1 5 10 15

Ala Leu Lys Lys Asn Glu Leu Leu Ser Glu Ala Gln Ser Ser Asp Ile
20 25 30

Leu Val Trp Asn Leu Thr Pro Arg Gln Leu Cys Asp Ile Glu Leu Ile
35 40 45

Leu Asn Gly Gly Phe Ser Pro Leu Thr Gly Phe Leu Asn Glu Asn Asp
50 55 60

Tyr Ser Ser Val Val Thr Asp Ser Arg Leu Ala Asp Gly Thr Leu Trp
65 70 75 80

Thr Ile Pro Ile Thr Leu Asp Val Asp Glu Ala Phe Ala Asn Gln Ile
85 90 95

```
Lys Pro Asp Thr Arg Ile Ala Leu Phe Gln Asp Asp Glu Ile Pro Ile
                100                 105                 110

Ala Ile Leu Thr Val Gln Asp Val Tyr Lys Pro Asn Lys Thr Ile Glu
        115                 120                 125

Ala Glu Lys Val Phe Arg Gly Asp Pro Glu His Pro Ala Ile Ser Tyr
    130                 135                 140

Leu Phe Asn Val Ala Gly Asp Tyr Tyr Val Gly Gly Ser Leu Glu Ala
145                 150                 155                 160

Ile Gln Leu Pro Gln His Tyr Asp Tyr Pro Gly Leu Arg Lys Thr Pro
                165                 170                 175

Ala Gln Leu Arg Leu Glu Phe Gln Ser Arg Gln Trp Asp Arg Val Val
            180                 185                 190

Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala His Arg Glu Leu Thr
        195                 200                 205

Val Arg Ala Ala Arg Glu Ala Asn Ala Lys Val Leu Ile His Pro Val
    210                 215                 220

Val Gly Leu Thr Lys Pro Gly Asp Ile Asp His His Thr Arg Val Arg
225                 230                 235                 240

Val Tyr Gln Glu Ile Ile Lys Arg Tyr Pro Asn Gly Ile Ala Phe Leu
                245                 250                 255

Ser Leu Leu Pro Leu Ala Met Arg Met Ser Gly Asp Arg Glu Ala Val
            260                 265                 270

Trp His Ala Ile Ile Arg Lys Asn Tyr Gly Ala Ser His Phe Ile Val
        275                 280                 285

Gly Arg Asp His Ala Gly Pro Gly Lys Asn Ser Lys Gly Val Asp Phe
    290                 295                 300

Tyr Gly Pro Tyr Asp Ala Gln Glu Leu Val Glu Ser Tyr Lys His Glu
305                 310                 315                 320

Leu Asp Ile Glu Val Val Pro Phe Arg Met Val Thr Tyr Leu Pro Asp
                325                 330                 335

Glu Asp Arg Tyr Ala Pro Ile Asp Gln Ile Asp Thr Thr Lys Thr Arg
            340                 345                 350

Thr Leu Asn Ile Ser Gly Thr Glu Leu Arg Arg Arg Leu Arg Val Gly
        355                 360                 365

Gly Glu Ile Pro Glu Trp Phe Ser Tyr Pro Glu Val Val Lys Ile Leu
    370                 375                 380

Arg Glu Ser Asn Pro Pro Arg Pro Lys Gln Gly Phe Ser Ile Val Leu
385                 390                 395                 400

Gly Asn Ser Leu Thr Val Ser Arg Glu Gln Leu Ser Ile Ala Leu Leu
                405                 410                 415

Ser Thr Phe Leu Gln Phe Gly Gly Gly Arg Tyr Tyr Lys Ile Phe Glu
            420                 425                 430

His Asn Asn Lys Thr Glu Leu Leu Ser Leu Ile Gln Asp Phe Ile Gly
        435                 440                 445

Ser Gly Ser Gly Leu Ile Ile Pro Asn Gln Trp Glu Asp Asp Lys Asp
    450                 455                 460

Ser Val Val Gly Lys Gln Asn Val Tyr Leu Leu Asp Thr Ser Ser Ser
```

```
465                 470                 475                 480

Ala Asp Ile Gln Leu Glu Ser Ala Asp Glu Pro Ile Ser His Ile Val
                485                 490                 495

Gln Lys Val Val Leu Phe Leu Glu Asp Asn Gly Phe Phe Val Phe
            500                 505                 510


<210> 31
<211> 309
<212> PRT
<213> Thermomonospora fusca

<400> 31
Met Ser Gln Val Ser Asp Ala Val Gly Arg Tyr Gln Leu Ser Gln Leu
  1               5                  10                  15

Asp Phe Leu Glu Ala Glu Ala Ile Phe Ile Met Arg Glu Val Ala Ala
             20                  25                  30

Glu Phe Glu Arg Pro Val Leu Leu Phe Ser Gly Gly Lys Asp Ser Val
         35                  40                  45

Val Met Leu Arg Ile Ala Glu Lys Ala Phe Trp Pro Ala Pro Ile Pro
     50                  55                  60

Phe Pro Val Met His Val Asp Thr Gly His Asn Phe Pro Glu Val Ile
 65                  70                  75                  80

Glu Phe Arg Asp Lys Arg Val Ala Glu Leu Gly Val Arg Leu Ile Val
                 85                  90                  95

Ala Ser Val Gln Asp Leu Ile Asp Ala Gly Lys Val Val Glu Pro Lys
            100                 105                 110

Gly Arg Trp Ala Ser Arg Asn Arg Leu Gln Thr Ala Ala Leu Leu Glu
        115                 120                 125

Ala Ile Glu Lys Tyr Gly Phe Asp Ala Ala Phe Gly Gly Ala Arg Arg
    130                 135                 140

Asp Glu Glu Lys Ala Arg Ala Lys Glu Arg Val Phe Ser Phe Arg Asp
145                 150                 155                 160

Glu Phe Gly Gln Trp Asp Pro Lys Asn Gln Arg Pro Glu Leu Trp Asn
                165                 170                 175

Leu Tyr Asn Thr Arg Val His Arg Gly Glu Asn Ile Arg Val Phe Pro
            180                 185                 190

Leu Ser Asn Trp Thr Glu Leu Asp Val Trp His Tyr Ile Arg Arg Glu
        195                 200                 205

Gly Leu Arg Leu Pro Ser Ile Tyr Phe Ala His Arg Arg Arg Val Phe
    210                 215                 220

Glu Arg Asp Gly Ile Leu Leu Pro Asp Ser Pro Tyr Val Thr Arg Asp
225                 230                 235                 240

Glu Asp Glu Glu Val Phe Glu Ala Ser Val Arg Tyr Arg Thr Val Gly
                245                 250                 255

Asp Met Thr Cys Thr Gly Ala Val Leu Ser Thr Ala Thr Thr Leu Asp
            260                 265                 270

Glu Val Ile Ala Glu Ile Ala Ala Thr Arg Ile Thr Glu Arg Gly Gln
        275                 280                 285

Thr Arg Ala Asp Asp Arg Gly Ser Glu Ala Ala Met Glu Glu Arg Lys
    290                 295                 300
```

Arg Glu Gly Tyr Phe
305

Claims

1. A polypeptide comprising a thermostable sulfurylase polypeptide comprising:

   a) an amino acid sequence of SEQ ID NO:2;
   b) a mature form of an amino acid sequence of SEQ ID NO:2;
   c) a variant of a mature form of an amino acid sequence of SEQ ID NO:2;
   d) a variant of an amino acid sequence of SEQ ID NO:2 wherein said variant differs in no more than 5% of amino acid residues; or
   e) a polypeptide having at least one conservative amino acid substitution to the amino acid sequences in a), b), c) or d).

2. A polypeptide according to claim 1, comprising SEQ ID NO:2.

3. A polypeptide according to claim 1 or 2, wherein the polypeptide further comprises a hexahistidine tag and a BCCP tag and has the amino acid sequence of SEQ ID NO:6.

4. A polypeptide according to claim 3, further comprising a luciferase polypeptide and having the sequence of SEQ ID NO:4.

5. A polypeptide according to any one of claims 1 to 4, wherein the polypeptide is attached to a mobile support.

6. A polypeptide according to claim 5, wherein the polypeptide is attached to the mobile support by a covalent or non-covalent interaction.

7. A polypeptide according to claim 6, wherein the polypeptide is attached by linkage selected from the group consisting of a Co^{2+} hexahistidine linkage, a Ni^{2+} hexahistidine linkage, a biotin binding protein/biotin linkage, a glutathione S-transferase/glutathione linkage, a monoclonal antibody/antigen linkage, a maltose binding protein/maltose linkage and a pluronic linkage.

8. A polypeptide according to claim 7, wherein the biotin binding protein is selected from the group consisting of streptavidin and avidin.

9. A polypeptide according to claim 5, wherein the mobile support is selected from the group consisting of a bead, optical fibre, and glass surface.

10. A polypeptide according to claim 9, wherein the bead is a nickel-agarose bead or a magnetic porous glass streptavidin bead.

11. An isolated nucleic acid comprising a sequence encoding a thermostable sulfurylase comprising:

    a) a nucleic acid encoding a sulfurylase polypeptide according to claim 1;
    b) a nucleotide sequence encoding a protein wherein the protein comprises an amino acid sequence at least 96% to mologous to the amino acid sequence of SEQ ID NO:2;
    c) a nucleic acid of SEQ ID NO:1;
    d) a nucleic acid of SEQ ID NO:1 in which no more than 11% of the nucleotides in the coding sequence differ from the coding sequence of SEQ ID NO:1.

12. An isolated nucleic acid according to claim 11, comprising a nucleotide sequence of SEQ ID NO:1.

13. An isolated nucleic acid according to claim 11 or 12, wherein the thermostable sulfurylase further comprises a hexahistidine tag and BCCP tag and has the sequence of SEQ ID NO:5.

14. An isolated nucleic acid according to claim 13, wherein the encoded polypeptide further comprises a luciferase polypeptide, having the sequence of SEQ ID NO:3.

15. An expression vector comprising a nucleic acid molecule of according to any one of claims 11 to 14.

16. A transformed host cell which contains the expression vector of claim 15.

17. A transformed host cell according to claim 16 wherein the host cell is a eukaryotic cell, optionally wherein the eukaryotic cell is a human, rat or mouse cell.

18. A transformed host cell according to claim 16 wherein the host cell is a prokaryotic cell, optionally wherein the prokaryotic cell is a bacterial cell.

FIGURE 1

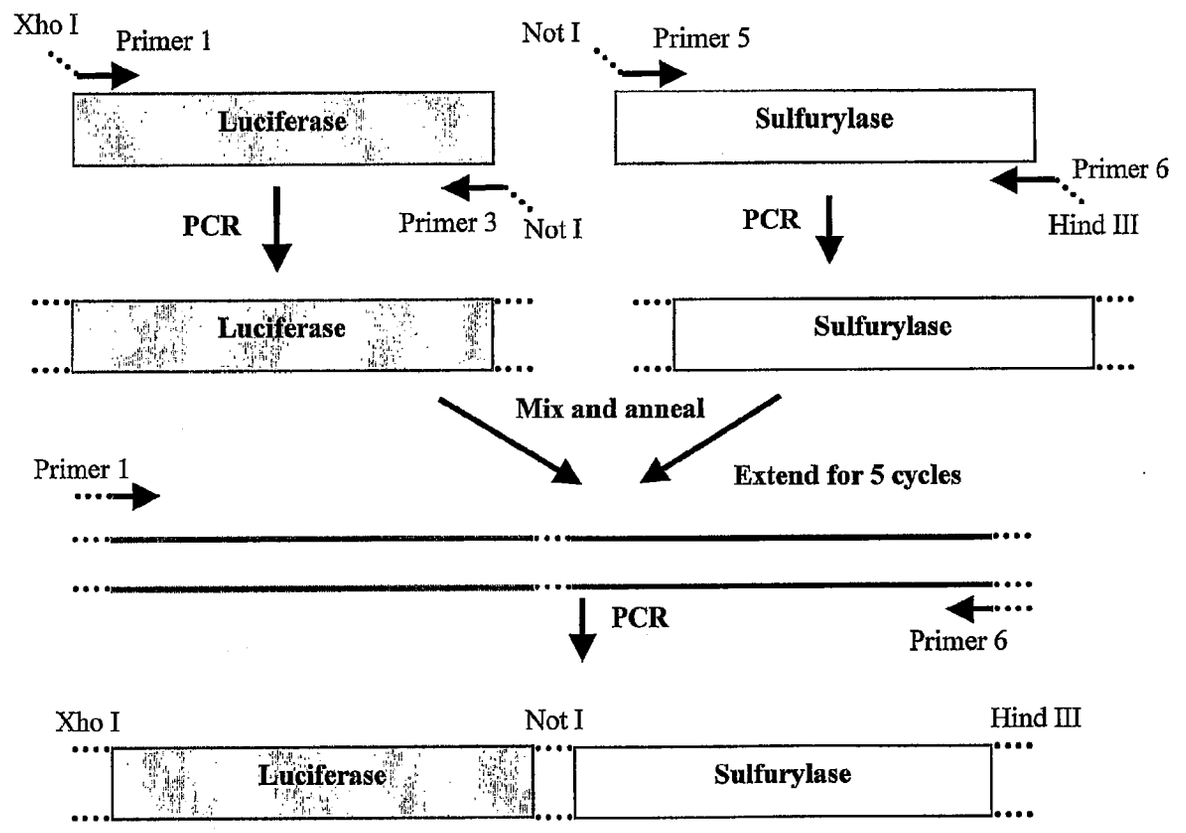
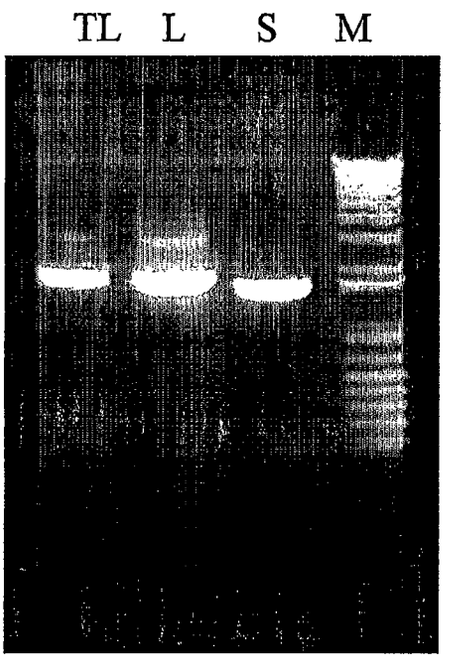
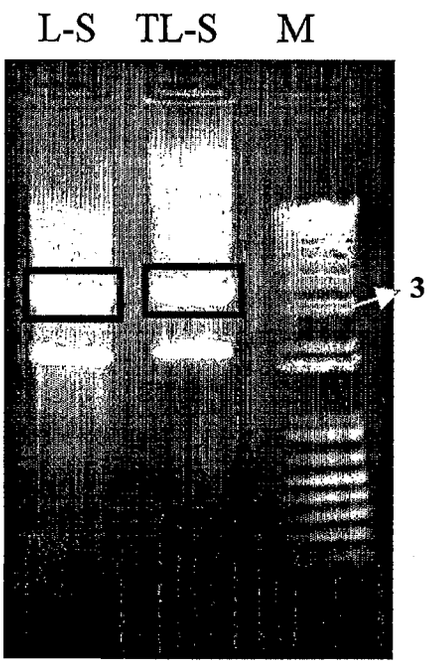

Xho I
Primer 1
Not I
Primer 5
Luciferase
Sulfurylase
Primer 6
PCR
Primer 3
Not I
PCR
Hind III
Luciferase
Sulfurylase
Mix and anneal
Primer 1
Extend for 5 cycles
PCR
Primer 6
Xho I
Not I
Hind III
Luciferase
Sulfurylase

Figure 2

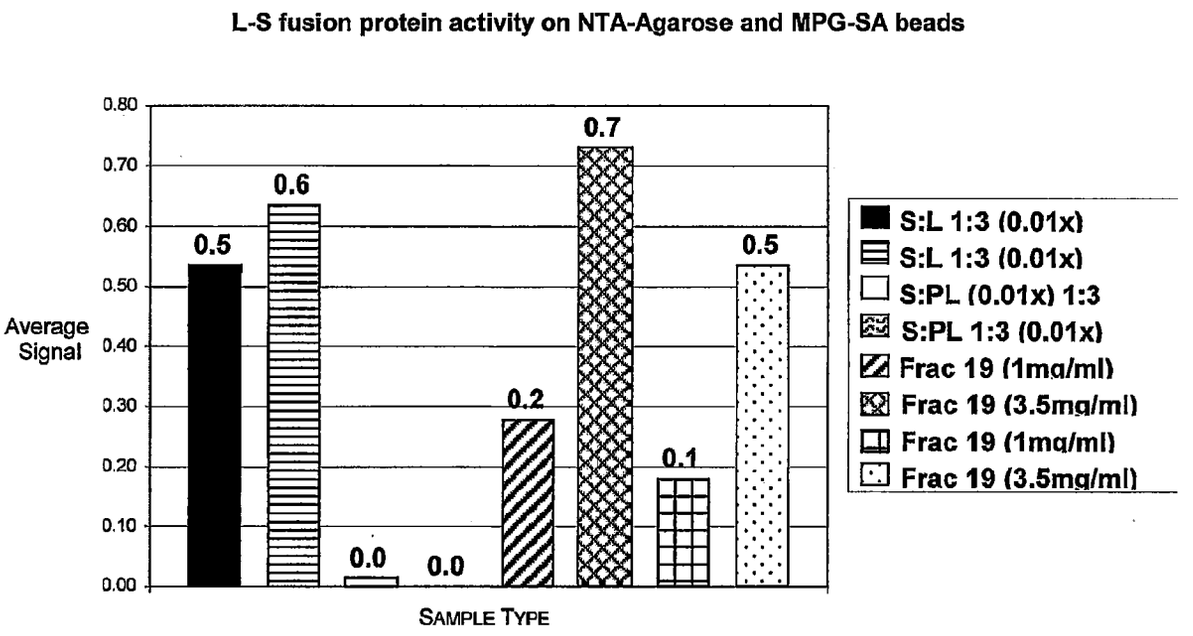


2A



2B

FIGURE 3


L-S fusion protein activity on NTA-Agarose and MPG-SA beads
Average Signal
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.5
0.6
0.0
0.0
0.2
0.7
0.1
0.5
SAMPLE TYPE
S:L 1:3 (0.01x)
S:L 1:3 (0.01x)
S:PL (0.01x) 1:3
S:PL 1:3 (0.01x)
Frac 19 (1mg/ml)
Frac 19 (3.5mg/ml)
Frac 19 (1mg/ml)
Frac 19 (3.5mg/ml)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 10 18 4482

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	ONDA M ET AL: "PURIFICATION AND PROPERTIES OF ADENOSINE 5'-TRIPHOSPHATE SULFURYLASE FROM THE THERMOPHILIC BACTERIUM BACILLUS STEAROTHERMOPHILUS", BIOSCIENCE BIOTECHNOLOGY BIOCHEMISTRY, vol. 60, no. 10, 1996, pages 1740-1742, XP009005919, JAPAN SOC. FOR BIOSCIENCE, BIOTECHNOLOGY AND AGROCHEM, TOKYO, JP ISSN: 0916-8451	1-3, 5-13, 15-18	INV. C12N1/21 C12N9/12 C12N15/52 C12N15/62
Y	* the whole document *	4,14	
X	DATABASE UniProt [Online] 1 October 2000 (2000-10-01), "Sulfate adenylyltransferase.", XP002367600, retrieved from EBI accession no. UNIPROT:Q9KCT2 Database accession no. Q9KCT2	1-3, 5-13, 15-18	
Y	* abstract *	4,14	TECHNICAL FIELDS SEARCHED (IPC)
Y	NYREN P ET AL: "Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis", ANALYTICAL BIOCHEMISTRY, vol. 151, no. 2, 1 December 1985 (1985-12-01), pages 504-509, XP024829646, ACADEMIC PRESS INC, NEW YORK ISSN: 0003-2697, DOI: 10.1016/0003-2697(85)90211-8 [retrieved on 1985-12-01] * the whole document *	4,14	C12N

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	29 March 2011	Weikl, Martina

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 10 18 4482

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
Y	CAPALDI S ET AL: "Signal amplification through nucleotide extension and excision on a dendritic DNA platform.", NUCLEIC ACIDS RESEARCH, vol. 28, no. 7, E21, 1 April 2000 (2000-04-01), pages I-VIII, XP002626366, ISSN: 1362-4962 * page ii, column 2, paragraph 3 - page iii, column 1, paragraph 1 *	4,14	
Y	MATSUNAGA T ET AL: "Production of luciferase-magnetic particle complex by recombinant Magnetospirillum sp. AMB-1", BIOTECHNOLOGY AND BIOENGINEERING, vol. 70, no. 6, 20 December 2000 (2000-12-20), pages 704-709, XP002367587, ISSN: 0006-3592 * the whole document *	4,14	
A	US 6 265 177 B1 (SQUIRRELL DAVID JAMES [GB] ET AL) 24 July 2001 (2001-07-24) * sequence 1 *	4,14	TECHNICAL FIELDS SEARCHED (IPC)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	29 March 2011	Weikl, Martina

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 10 18 4482

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

29-03-2011

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6265177 B1	24-07-2001	AT 233810 T	15-03-2003
		AU 731408 B2	29-03-2001
		AU 6929398 A	11-11-1998
		CA 2283287 A1	22-10-1998
		DE 69811876 D1	10-04-2003
		DE 69811876 T2	04-12-2003
		DK 973874 T3	19-05-2003
		EP 0973874 A2	26-01-2000
		ES 2194309 T3	16-11-2003
		WO 9846729 A2	22-10-1998
		JP 4302777 B2	29-07-2009
		JP 2001518799 T	16-10-2001
		NZ 337373 A	27-04-2001
		PT 973874 E	31-07-2003

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 4873316 A **[0070]**
- EP 264166 A **[0070]**
- WO 9813523 A1 **[0080]**
- US 6274320 B **[0080]**
- US SN10104280 A **[0080]**
- US 5854033 A **[0094] [0103] [0107]**
- US 5714320 A, Kool **[0103] [0107]**

Non-patent literature cited in the description

- **ROBBINS ; LIPMANN.** *J. Biol. Chem.,* 1958, vol. 233, 686-690 **[0002]**
- **HAWES ; NICHOLAS.** *Biochem. J.,* 1973, vol. 133, 541-550 **[0002]**
- MOLECULAR CLONING: A LABORATORY MANUAL. Cold Spring Harbor Laboratory Press, 1989 **[0043]**
- CURRENT PROTOCOLS IN MOLECULAR BIOLOGY. John Wiley & Sons, 1993 **[0043]**
- **AUSUBEL et al.** CURRENT PROTOCOLS IN MOLECULAR BIOLOGY. John Wiley & Sons, 1993 **[0047]**
- **WILSON ; BANDURSKI.** *J. Biol. Chem.,* 1958, vol. 233, 975-981 **[0050]**
- **SEUBERT et al.** *Arch. Biochem. Biophys.,* 1983, vol. 225, 679-691 **[0050]**
- **SEUBERT et al.** *Arch. Biochem. Biophys.,* 1985, vol. 240, 509-523 **[0050]**
- **DE WET et al.** *Proc. Natl. Acad. Sci. USA,* 1985, vol. 80, 7870-7873 **[0051]**
- **OW et al.** *Science,* 1986, vol. 234, 856-859 **[0051]**
- **JHA et al.** *FEBS Lett.,* 1990, vol. 274, 24-26 **[0051]**
- **DE WET et al.** *Mol. Cell. Biol.,* 1987, vol. 7, 725-7373 **[0051]**
- **KELLER et al.** *Proc. Natl. Acad. Sci. USA,* 1987, vol. 82, 3264-3268 **[0051]**
- **WOOD et al.** *J. Biolumin. Chemilumin.,* 1989, vol. 4, 289-301 **[0051]**
- **WOOD et al.** *Science,* 1989, vol. 244, 700-702 **[0051]**
- **MCELROY ; SELINGER.** *Arch. Biochem. Biophys.,* 1960, vol. 88, 136-145 **[0052]**
- **LEACH.** *J. Appl. Biochem.,* 1981, vol. 3, 473-517 **[0052]**
- **GREEN ; KRICKA.** *Talanta,* 1984, vol. 31, 173-176 **[0052]**
- **BLUM et al.** *J. Biolumin. Chemilumin,* 1989, vol. 4, 543-550 **[0052]**
- Applications of firefly luciferase. **LUNDIN.** Luminescent Assays. Raven Press, 1982 **[0053]**
- **NILSSON et al.** *EMBO J.,* 1985, vol. 4, 1075 **[0054]**
- **NILSSON et al.** *Methods Enzymol.,* 1991, vol. 198, 3 **[0054]**
- **SMITH ; JOHNSON.** *Gene,* 1988, vol. 67, 31 **[0054]**
- **HOPP et al.** Biotechnology. Eastman Kodak Co, 1988, vol. 6, 1204-1210 **[0054]**
- **FORD et al.** *Protein Expression and Purification,* 1991, vol. 2, 95-107 **[0054]**
- CURRENT PROTOCOLS IN MOLECULAR BIOLOGY. John Wiley & Sons, 1992 **[0057]**
- **GOEDDEL.** GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY. Academic Press, 1990, vol. 185 **[0064] [0065]**
- **BALDARI et al.** *EMBO J,* 1987, vol. 6, 229-234 **[0067]**
- **KURJAN ; HERSKOWITZ.** *Cell,* 1982, vol. 30, 933-943 **[0067]**
- **SCHULTZ et al.** *Gene,* 1987, vol. 54, 113-123 **[0067]**
- **SMITH et al.** *Mol Cell Biol,* 1983, vol. 3, 2156-2165 **[0068]**
- **LUCKLOW ; SUMMERS.** *Virology,* 1989, vol. 170, 31-39 **[0068]**
- **SEED.** *Nature,* 1987, vol. 329, 840 **[0069]**
- **KAUFMAN et al.** *EMBO J,* 1987, vol. 6, 187-195 **[0069]**
- **SAMBROOK et al.** MOLECULAR CLONING: A LABORATORY MANUAL. Cold Spring Harbor Laboratory Press, 1989 **[0069] [0073]**
- **PINKERT et al.** *Genes Dev,* 1987, vol. 1, 268-277 **[0070]**
- **CALAME ; EATON.** *Adv Immunol,* 1988, vol. 43, 235-275 **[0070]**
- **WINOTO ; BALTIMORE.** *EMBO J,* 1989, vol. 8, 729-733 **[0070]**
- **BANERJI et al.** *Cell,* 1983, vol. 33, 729-740 **[0070]**
- **QUEEN ; BALTIMORE.** *Cell,* 1983, vol. 33, 741-748 **[0070]**
- **BYRNE ; RUDDLE.** *PNAS,* 1989, vol. 86, 5473-5477 **[0070]**
- **EDLUND et al.** *Science,* 1985, vol. 230, 912-916 **[0070]**

- **KESSEL ; GRUSS.** *Science,* 1990, vol. 249, 374-379 **[0070]**
- **CAMPES ; TILGHMAN.** *Genes Dev,* 1989, vol. 3, 537-546 **[0070]**
- *Journal of Biological Chemistry,* 1988, vol. 263, 6461 **[0077]**
- **HENGSAKUL ; CASS.** *Bioconjugate Chem.,* 1996, vol. 7, 249-254 **[0077]**
- **PIRRUNG ; HUANG.** *Bioconjugate Chem.,* 1986, vol. 7, 317-321 **[0078]**
- **RONAGHI et al.** *Anal. Biochem.,* 1996, vol. 242, 84-89 **[0080]**
- **RONAGHI et al.** *Science,* 1998, vol. 281, 363-365 **[0080]**
- **NILSSON et al.** *Science,* 1994, vol. 265, 2085-2088 **[0095] [0111]**
- **LIZARDI et al.** *Nat. Genet.,* 1998, vol. 19, 225-232 **[0095] [0109]**
- **NILSSON et al.** *Nat. Genet.,* 1997, vol. 16, 252-255 **[0095]**
- **LANDEGREN ; NILSSON.** *Ann. Med.,* 1997, vol. 29, 585-590 **[0095]**
- **HATCH et al.** *Genet. Anal. Biomol. Engineer.,* 1999, vol. 15, 35-40 **[0103] [0111]**
- **SAIKI et al.** *Science,* 1995, vol. 230, 1350-1354 **[0106]**
- **BARANY.** *Proc. Natl. Acad. Sci. USA,* 1991, vol. 88, 189-193 **[0106]**
- **BARRINGER et al.** *Gene,* 1990, vol. 89, 117-122 **[0106]**
- **KWOH et al.** *Proc. Natl. Acad. Sci. USA,* 1989, vol. 86, 1173-1177 **[0106]**
- **GUATELLI et al.** *Proc. Natl. Acad. Sci. USA,* 1990, vol. 87, 1874-1878 **[0106] [0113]**
- **LIZARDI et al.** *BioTechnology,* 1988, vol. 6, 1197-1202 **[0106] [0113]**
- *Nucleic Acids Res.,* 11 April 1992, vol. 20 (7), 1691-6 **[0106]**
- *PNAS,* 01 January 1992, vol. 89 (1), 392-6 **[0106]**
- *NASBA J Virol Methods,* December 1991, vol. 35 (3), 273-86 **[0106]**
- **HATCH et al.** *Genet. Anal. Biomol. Engineer,* 1999, vol. 15, 35-40 **[0107]**
- **ZHANG et al.** *Gene,* 1998, vol. 211, 277-85 **[0111]**
- **BANER et al.** *Nucl. Acids Res.,* 1998, vol. 26, 5073-5078 **[0111]**
- **LIU et al.** *J. Am. Chem. Soc.,* 1995, vol. 118, 1587-1594 **[0111]**
- **FIRE ; XU.** *Proc. Natl. Acad. Sci. USA,* 1995, vol. 92, 4641-4645 **[0111]**
- **KIEVITS et al.** *J. Virol. Methods,* 1991, vol. 35, 273-286 **[0113]**